**UNIVERSIDAD NACIONAL DE GENERAL SAN MARTÍN**
**COMISIÓN NACIONAL DE ENERGÍA ATÓMICA**
**INSTITUTO DE TECNOLOGÍA**
**"Prof. Jorge A. Sabato"**

# Fenomenología en búsquedas de Nueva Física pesada y cuatro-tops en el LHC [(*)]

## por Lic. Manuel Szewc

**Director**

**Dr. Ezequiel Alvarez**

**Codirector**

**Dr. Alejandro Szynkman**

**[(*)] Tesis para optar al título de Doctor en Ciencia y Tecnología, mención Física**

**República Argentina**

**2021**

**UNIVERSIDAD NACIONAL DE GENERAL SAN MARTÍN**
**COMISIÓN NACIONAL DE ENERGÍA ATÓMICA**
**INSTITUTO DE TECNOLOGÍA**
**"Prof. Jorge A. Sabato"**

# Phenomenology in four-top and heavy New Physics searches at the LHC (*)

**Manuel Szewc**

**Supervisor**

**Dr. Ezequiel Alvarez**

**Co-supervisor**

**Dr. Alejandro Szynkman**

**(*) Thesis submitted for the degree of Doctor in Science and Technology, Physics**

**República Argentina**

**2021**

Esta Tesis está dedicada a mi familia.

A mis viejos, a mis hermanos y a mis suegros.

Y, especialmente, a la única persona que me sigue eligiendo todos los días.

# RESUMEN

En esta Tesis consideramos diferentes estrategias para detectar efectos de Física Más Allá del Modelo Estándar (BSM) en el LHC. Estas estrategias pueden dividirse en dos tipos: búsquedas con modelos específicos y búsquedas agnósticas al modelo. Las primeras se concentran en (una familia de) modelos de BSM cuya fenomenología única provee rastros interesantes en el LHC; mientras que las segundas apuntan a desarrollar análisis semi- y no-supervisados que encuentren rastros de efectos de BSM. Estas dos estrategias son complementarias. Aunque las búsquedas agnósticas al modelo son útiles para encontrar la aguja en un pajar, bajo ciertas hipótesis, y evitar potenciales errores sistemáticos provenientes de las simulaciones de Monte Carlo, deben realizarse búsquedas dedicadas para aumentar el poder de discriminacion y extraer parámetros fisicamente relevantes, como la sección eficaz y los acoples, con la ayuda de simulaciones de Monte Carlo.

Entre los muchos escenarios posibles de BSM que pueden ser explorados por el LHC, consideramos tres posibilidades. Cada una de estas posibilidades está motivada por diferentes resultados experimentales y da lugar a diferentes estados finales, con cierta superposición entre los tres casos. Primero exploramos un Leptoquark inspirado por las anomalías en la Física de mesones B. Este Leptoquark es un bosón de seudo Nambu-Goldstone de una teoría de campos fuertemente acoplada y es producido de a pares en el LHC. Luego, consideramos como un Leptoquark multigeneracional que puede ser producido de manera no resonante en el LHC puede evitar límites existentes debido a dicho comportamiento no resonante. Finalmente, mostramos como un FCNC $Z'$ interactuando con el quark top podría estar presente en estados finales multileptónicos con alta multiplicidad de jets b, explicando discrepancias persistentes en los análisis de $t\bar{t}W^{\pm}$ y 4-top. En todos los casos, reutilizamos búsquedas existentes para establecer límites en los espacios de parámetros de nuestro modelos y proponemos nuevas búsquedas que excluirian regiones significativas de dichos espacios de parámetros.

Para implementar búsquedas agnósticas al modelo, nos concentramos en dos algoritmos de modelado de tópicos que adaptamos a la física del LHC. Estudiamos cómo el algoritmo conocido como `Demixer`, previamente aplicado a clasificación de quarks y gluones, puede detectar mezclas estadísticas de 4-top y $t\bar{t}W^{\pm} + t\bar{t}H$ en el estado final dileptónico y provee un método para adecuar las simulaciones de Monte Carlo de la distribución de $t\bar{t}W^{\pm} + t\bar{t}H$ utilizando las mismas regiones de señal analizadas. Luego mostramos cómo se puede utilizar el algoritmo conocido como Latent Dirichlet Allocation para realizar búsquedas no-supervisadas en el LHC. Estudiamos en el espacio de hiperparametros y mostramos la correlación entre la Perplejidad y la capacidad de clasificación, obteniendo una métrica no-supervisada para seleccionar hiperparametros.

Tras haber explorado distintos análisis basados tanto en simulaciones de Monte Carlo como en algoritmos no supervisados, encontramos que estos últimos parecen ser prometedores, con mucho margen de mejora en los próximos años.

# ABSTRACT

In this Thesis we consider different strategies for detecting Beyond the Standard Model (BSM) effects at the LHC. These strategies can be split in two sets: model specific and model agnostic searches. The former studies target (a family of) BSM models whose unique phenomenology provides interesting fingerprints at the LHC while the latter aim to develop semi-supervised or unsupervised analyses that point towards BSM effects. These two strategies are complementary. Although model agnostic searches are useful to find the needle in the haystack (under certain assumptions) and avoid potential systematic errors originating in the Monte Carlo simulations, dedicated searches should be performed afterwards to increase the discriminating power and extract physically relevant parameters, such as cross-sections and couplings, with the help of Monte Carlo simulations.

Among the many BSM scenarios that could be probed by the LHC, we consider three possibilities. Each of these possibilities is motivated by different experimental results and have different, overlapping, final states. We first explore a B-Anomalies-inspired Leptoquark arising as a pseudo Nambu-Goldstone boson of a strongly coupled field theory that is pair produced at the LHC. Next, we consider how a multi-generational Leptoquark that can be produced non-resonantly at the LHC could avoid existing constraints by virtue of said non-resonant behavior. Finally, we show how a FCNC top-philic $Z'$ that could be present in multilepton with high b-jet multiplicity final states and explain recent persistent discrepancies in $t\bar{t}W^{\pm}$ and 4-top analyses. In all cases, we recast existing searches to constrain these models' parameter spaces and we propose new searches that should be able to exclude a significant region of said parameter spaces.

To implement model agnostic searches, we focus on two topic modeling algorithms adapted to LHC physics. We study how the `Demixer` algorithm previously applied to quark/gluon tagging could detect statistical mixtures of 4-top and $t\bar{t}W^{\pm} + t\bar{t}H$ at the dilepton final state. We observe that this algorithm provides a method to tune Monte Carlo simulations of background distribution in the signal regions. We also show how Latent Dirichlet Allocation can be used to perform unsupervised searches at the LHC. We perform hyperparameter scans and demonstrate the correlation between Perplexity and tagging performance, thus obtaining an unsupervised metric for hyperparameter selection.

Having explored different Monte Carlo-based and unsupervised analyses to search for BSM physics, we find that the latter seems a fertile ground with much room for improvement over the coming years.

# AGRADECIMIENTOS

Esta tesis es la culminación de tres años de trabajo en los que tuve la oportunidad de interactuar con y aprender de una gran cantidad de personas. Todos los trabajos que componen esta tesis son un reflejo de estas interacciones.

La interacción más fuerte, y a quien le quiero agradecer en primer lugar, es a mi director, Ezequiel Alvarez. Tuve la suerte de poder aprender de él (y con él) no sólo de física e inteligencia artificial sino que además incorpore una manera de trabajar y de entender la vocación por la ciencia.

Quiero agradecer también a mi co-director, Alejandro Szynkman, que ha sido la voz de la consciencia durante este tiempo. Siempre que quiero estar seguro de lo que hago, pienso en lo que preguntaría Alejandro.

Además de Ezequiel y Alejandro, tuve el privilegio de poder trabajar directamente con mucha gente a la que le quiero agradecer: Leandro Da Rold, Tamara Vazquez Schroeder, Barry Dillon, Dario Faroughy, Aurelio Juste y Jernej Kamenik. En particular, les estoy muy agradecido a Aurelio y Jernej por recibirme con los brazos abiertos. Otro colaborador al que le guardo particular aprecio es a Federico Lamagna, con quien comparti más momentos de los que entran aca.

Nada de esto sería posible sin el apoyo del ICAS. En particular, le estoy muy agradecido a Daniel de Florian, quien siempre me ayudó enormemente, y a Rodrigo Diaz, con quien tengo el privilegio de dar clases. También le quiero agradecer a todos los becarios con quien tuve el lujo de compartir oficina: Rosa, Mariel, Nacho Fabre, Nacho Borsa, Manu, Ivan, Yami, Rocio, Leo, Pablo y Daniel. Por último, le quiero agradecer a Joaquin, por razones que él sabe mejor que nadie.

No me gustaría terminar esta sección sin agradecerle a todas las personas fuera del ámbito académico que me apoyaron incondicionalmente durante este tiempo. Lamentablemente, va a tener que ser así por cuestiones de espacio. Solo quiero agradecer a la persona que me convenció, en estos tiempos inciertos, de que se puede creer en un final feliz. Gracias Juan Roman Riquelme por volver a tu casa.

# CONTENTS

# 1 INTRODUCTION

The goal of this Ph.D. Thesis is, broadly speaking, to study the possibility of New Physics at the Large Hadron Collider experiment. As there is not a single more compelling extension of the Standard Model we approach this problem from a phenomenological point of view, subscribing to the idea that the way to find the light switch in a dark room is to explore it even if you risk bumping into the furniture.

To find the "light switch", we take two different approaches. Firstly, we take aim of different specific Beyond the Standard Model scenarios, motivated by existing experimental results, and study their phenomenology at the ATLAS and CMS experiments. Secondly, we implement Machine Learning algorithms to devise search strategies that either reduce Monte Carlo dependency or avoid the introduction of specific New Physic hypotheses by performing unsupervised or semi-supervised searches.

In this Introduction we detail some of the concepts needed along this Thesis. We briefly introduce the Standard Model of Particle Physics in Section 1.1, with the need of going Beyond the Standard Model and the specific cases we consider detailed in Section 1.2. We give a general introduction to the Topic Modelling techniques for Machine Learning in Section 1.3.

## 1.1 THE STANDARD MODEL OF PARTICLE PHYSICS

The Standard Model of particle physics (SM) is a largely tested and successful theory which seeks to understand the fundamental components of the universe. The language of the SM is that of Quantum Field Theory (QFT), which combines quantum mechanics and special relativity. A complete introduction to QFT and to the SM goes beyond the scope of this Thesis. We give here a brief review and we refer the interested reader to one of the excellent available literature, e.g. Refs. [1–4]. For our purposes, it suffices to say that using the QFT language we can write the SM Lagrangian, containing all relevant degrees of freedom and parameters, and we can use it to understand and predict key experimental observables. The fundamental components of the SM are embedded in quantum fields that transform under unitary irreducible representations of the Poincare symmetry group, with the SM particles defined through the normal modes of these fields. The particle content of the SM is shown in Table 1.

The quarks and leptons, plus their corresponding antiparticles, are embedded in fermionic spin 1/2 matter fields. There are 3 generations of quarks, both up and down. The three flavours of up-type quarks are the up $u$, charm $c$ and top $t$. They exhibit a clear mass hierarchy with approximate masses (in natural units) 2 MeV, 1 GeV and 170 GeV respectively [5]. The down-type quarks are down $d$, strange $s$ and bottom $b$. They also exhibit a mass hierarchy while being lighter that their same generation up-type quark, having approximate masses of 4 MeV, 90 MeV, and 4 GeV. Similarly, there are three generations of leptons. The charged leptons are the electron $e$, muon $\mu$ and tau $\tau$ leptons with approximate masses of 411 keV, 101 MeV and 1.78 GeV. The neutral neutrinos also have three flavours, electron $\nu_e$, muon $\nu_\mu$ and tau $\nu_\tau$ neutrinos. Although we know at least two of them are massive, we will consider them massless as they behave as such for the experimental data we consider. At high enough energy

| | Particle | $SU(3)_C$ | $SU(2)_L$ | $U(1)_Y$ |
|---|---|---|---|---|
| Quarks | $Q_L = \left(u_L^i, d_R^i\right)^T$ | 3 | 2 | 1/6 |
| | $u_R^i$ | 3 | 1 | 2/3 |
| | $d_R^i$ | 3 | 1 | $-1/3$ |
| Leptons | $\ell_L = \left(\nu_L^j, e_R^j\right)^T$ | 1 | 2 | $-1/2$ |
| | $e_R^j$ | 1 | 1 | $-1$ |
| Gauge Bosons | $\gamma$ | 1 | 1 | 0 |
| | $(W^+, Z, W^-)^T$ | 1 | 3 | 0 |
| | $g$ | 8 | 1 | 0 |
| Higgs Boson | $\left(0, \frac{H+v}{\sqrt{2}}\right)^T$ | 1 | 2 | 1/2 |

Table 1: The SM particle content grouped labelled by the SM Lie Group symmetries. For $SU(3)_C$ and $SU(2)_L$ we write the irreducible representations to which the associated field belongs by its multiplicity and for $U(1)_Y$ we write the charge of the associated field under $U(1)_Y$. We use Ref. [4]'s convention, which yields $Q_{em} = T_3 + Y$. The flavour indices in the quark and lepton sectors go from 1 to 3. We do not include right neutrinos $\nu_R$ although the observed neutrino mass implies their existence.

scales that all fermions become massless, the fermionic fields become chiral and the number of different fermionic fields nearly doubles (as there are left- and right-handed fields for every fermionic except the three neutrinos and their kinematic Lagrangian becomes decoupled without a non-chiral mass term).

We encode the different interactions between matter as local gauge symmetries of the SM Lagrangian. A local gauge symmetry is achieved by ensuring that the fields are also irreducible representation of the corresponding Lie group. The SM has three local gauge symmetries and thus three Lie groups under which the fields transform: $SU(3)_C - SU(2)_L - U(1)_Y$. We can label the fields and their corresponding particles by the irreducible representations to which they belong, and so each particle (or multiplet of particles) has three charges: colour, weak isospin and hypercharge. These local gauge symmetries give place to bosonic (spin 1) gauge fields and the corresponding massless gauge particles. Gauge symmetry ensures the renormalizability of the SM Lagrangian and, with the introduction of the electroweak symmetry breaking, provides an accurate description of Nature.

The electroweak symmetry breaking is necessary to give mass to the quarks and leptons and to obtain massive weak bosons $W^\pm$, $Z$ (with approximate masses 80 and 91 GeV respectively) while keeping a massless photon $\gamma$. This is achieved by introducing a complex scalar field that is an $SU(2)_L$ doublet, the Higgs field. This Higgs field has an interaction potential which, in a process known as Spontaneous Symmetry Breaking, makes the vacuum expectation value (vev) of the scalar field not invariant under $SU(2)_L$ transformations. This breaks the $SU(2)_L - U(1)_Y$ sector to a $U(1)_{EM}$ while keeping the theory renormalizable. Due to the interactions between the Higgs field and the different matter and gauge fields, the necessary masses arise and are proportional to the Higgs vev $v$. This is known as the Higgs mechanism. The Higgs

field has a corresponding particle, the spin 0 Higgs boson H of approximate mass 125 GeV.

While the interactions between the Higgs and the rest of the Standard Model fields are determined by the gauge symmetries, the specific flavour structure of the couplings between the Higgs and the fermions, which we call the Yukawa matrices Y, is arbitrary. Although these matrices are diagonal on the mass basis, whose mass eigenstates are the ones we measure at the detector level, they need not be so in the flavour basis, which is more useful basis for interactions. This introduces different mass and flavour eigenstates, a distinction we must deal with. In the quark sector, it is useful to introduce the unitary CKM matrix $V_{CKM}$ such that

$$\begin{aligned} u^j_{\text{mass}} &= (U_u)^{ji} u^i_{\text{flavour}} \\ d^j_{\text{mass}} &= (U_d)^{ji} d^i_{\text{flavour}} \\ V_{CKM} &= U_u^\dagger \cdot U_d \end{aligned}$$

The CKM matrix is responsible for a wide variety of effects, including CP violation. At the Lagrangian level, the difference between mass and flavour basis appears only in the $\bar{u}^i \gamma^\mu d^j W^+_\mu$ interaction vertex, which is proportional to $(V_{CKM})^{ij}$. There is an analogous unitary matrix for the lepton sector, the PMNS matrix $U_{PMNS}$. However, it is usually summed out because we are agnostic to neutrino flavour at the detector level. In the following, we always work in the mass basis and we disregard the flavour-changing interactions when the CKM elements are small enough.

Among the plethora of observables that test the SM, we focus on collider physics. Collider experiments are controlled environments where we can test SM predictions by colliding particles (be them electrons or baryons such as proton, antiproton and lead) and seeing the different processes that take place. These processes can be related to S-Matrix elements of our theory. These S-Matrix elements can most often be calculated with an array of QFT tools and thus allow us to test the SM. Depending on the particles being collided and the energies involved, different physical processes can take place. We focus on two collider experiments: ATLAS and CMS.

### 1.1.1 *ATLAS and CMS Phenomenology*

Installed at the CERN's Large Hadron Collider (LHC), the ATLAS (short for A Toroidal LHC ApparatuS) and CMS (short for Compact Muon Solenoid) experiments collide bunches of protons with the highest center-of-mass energies ever achieved in a laboratory, reaching center-of-mass energies of $\sqrt{s} = 13$ TeV at the end of Run-2. Proton-proton collisions have the virtue of probing a vast array of energies through the distribution functions of the QCD partons. Due to this, the LHC is a "discovery machine" designed mainly to find the Higgs Boson and push our knowledge of the fundamental components of matter forward. To achieve this, the ATLAS and CMS experiments possess general-purpose detectors designed to detect the broadest possible range of signals.

The ATLAS and CMS collaborations are pushing forward our understanding of Nature. They have discovered the Higgs boson [6, 7] and started measuring its properties [8, 9]. They have provided more and more accurate tests of the SM, both of the QCD (see e.g. [10]) and Electroweak sectors (where they are able to beat in precision previous precision colliders such as LEP-2 when measuring the $W^\pm$ bosons coupling

to leptons [11]). They have also looked for Beyond the Standard Model (BSM) effects, as we detail in Section 1.2.

In order to achieve this, most current searches done at ATLAS and CMS rely at some point in the analysis in being able to reproduce faithfully enough complicated SM phenomena. As SM calculations can present a number of complications, ranging from numerical integration to theoretical problems due to non-perturbative effects, validated Monte Carlo techniques are necessary to obtain accurate predictions.

The pipeline we consider to obtain theoretical simulations of the expected event yields consists of three steps:

1. Parton level generation, either at Leading Order (LO) or at Next-to-Leading Order (NLO) using `MadGraph5_aMC@NLO` [12]. `MadGraph5_aMC@NLO` simulates hard processes and decays the heavy particles (either directly at the generation level or as an extra step with the help of `MadSpin` [13], the latter being the only option for spin-correlated decays at NLO). BSM models are implemented with the help of `FeynRules` [14]. The output is a `LHE` file [15] which is only physical for LO but must be matched to a parton shower for NLO processes.

2. Showering and hadronization using `Pythia 8` [16]. With the help of extensive validation and tuning to the experimental data, `Pythia 8` is able to simulate the very soft physics that accompanies the hard process, performing the matching of soft and hard physics. This is necessary if one wants to obtain physical kinematic distributions at NLO. The output is a `HepMC` file [17].

3. Detector simulation using `Delphes 3` [18]. The output of `Pythia 8` consists of the physical objects which could be measured. To simulate the detector response to these physical objects and obtain the measured objects, we implement `Delphes 3`, which needs to be tuned and validated to match the experimental analyses we want to reproduce. The output is a `ROOT` file [19].

Although we always consider this specific pipeline, there are alternatives. Hard processes matrix elements could be calculated differently using event generators such as `Sherpa 2.2` [20]. `Herwig 7` [21, 22] provides an alternative to `Pythia 8`, and the difference between the obtained event samples using both programs can be used to obtain the uncertainty on the choice of parton shower and hadronization parameters. Finally, detector response is more precisely modelled if one uses the `GEANT4` package [23] at the cost of more computational power and detailed knowledge of the detector.

In each Chapter of this Thesis we explore different combinations of final states chosen from the plethora of available states at the LHC. Of particular interest are those final states that involve the heaviest, and thus more unexplored, sectors of the SM: the Higgs and top quark sectors. In addition to testing SM predictions, we are interested in how current experimental results constrain different Beyond the Standard Model (BSM) models and how to improve upon them by proposing new search strategies. Chapters 2, 3 and 4 explore specific BSM models scenarios detailed in Section 1.2. Each of these Chapters explores different (and often overlapping) sets of final states.

The ATLAS and CMS analyses are highly sophisticated, implementing a wide array of statistical techniques designed to discriminate signal from background. However, to have an accurate prediction of the expected background and signal events they usually rely on Monte Carlo simulations. These techniques are computationally expensive and, although extensively validated, introduce systematic uncertainties which limit the performance of the analyses. In Chapters 5 and 6 we detail how the use of

topic modeling techniques detailed in Section 1.3 could be useful to reduce Monte Carlo dependency and even perform unsupervised searches where some systematic modeling uncertainties are greatly reduced.

Before we proceed, we detail the phenomenology of a final state of particular interest: four-top-quarks or 4-top for short. In detailing this we aim to give an example of what we mean by studying the phenomenology of a given final state while also seeing in first hand the possibility of BSM physics exploration and the need for reducing the dependence on Monte Carlo simulations.

*4-top*

Four-top-quarks production $t\bar{t}t\bar{t}$,or 4-top, is among the most challenging final states to reconstruct at the LHC. The Leading Order (LO) production diagrams are shown in Fig. 1. The heavy mass of the top quark implies that, although the production is favored by the gluons' parton distribution function (pdf), the cross-section is highly suppressed by kinematic requirements. Although the pure QCD $\mathcal{O}(\alpha_S^2)$ diagrams dominate the cross-section, supposedly subleading Electroweak diagrams of $\mathcal{O}(\alpha_S \alpha_{EW})$ are comparable due to the complex structure of the amplitudes. This is due to resonant $tt \to tt$ contributions. The situation is even more complicated at Next to Leading Order (NLO) where sophisticated calculations are involved to give an state of the art calculation of $\sigma_{t\bar{t}t\bar{t}} = 11.97^{+18\%}_{-21\%}$ where the theoretical uncertainties come from the cross-section dependence on the choice of factorization and renormalization scales [24].

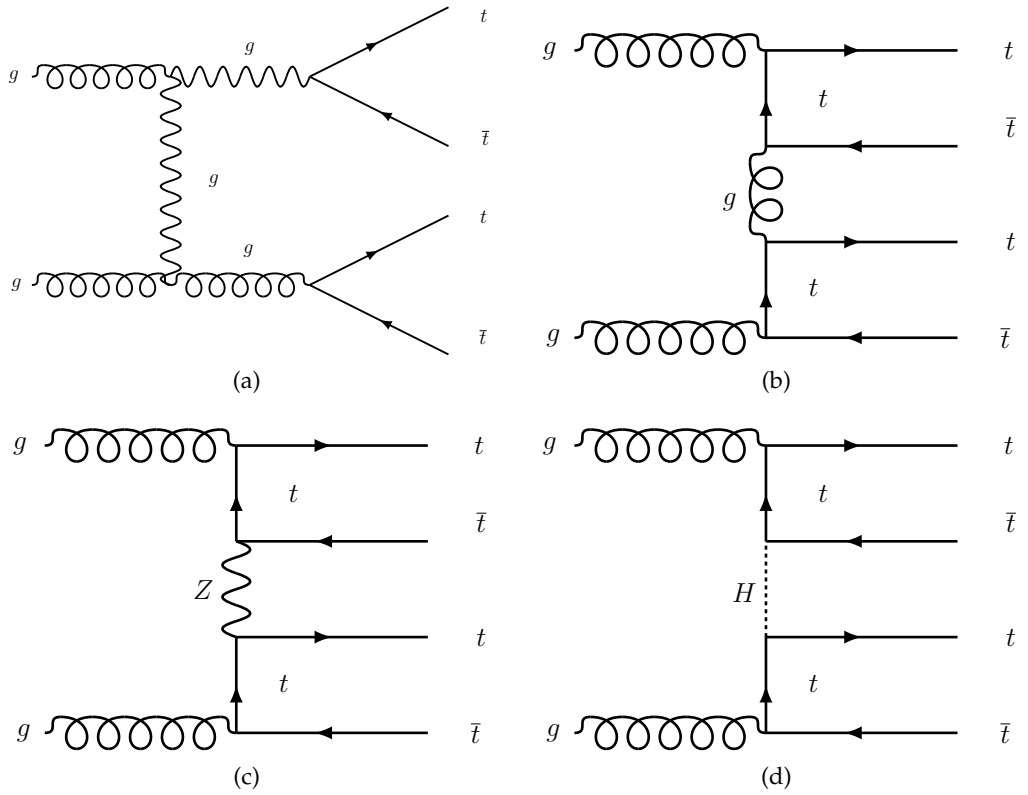



Figure 1: Representative diagrams for LO 4-top production.

From the experimental side, 4-top measurements are also very challenging. In addition to the small cross-section, the reconstruction of the final state is difficult because the four top quarks decay before reaching detection. The top quarks decay almost

exclusively to $Wb$. The $W$ boson has an approximate 67% Branching Ratio to light quarks, 22% Branching Ratio to electrons and muons and 11% to taus with their corresponding neutrinos. The latter also decay before reaching the detector, having approximately a 65% Branching Ratio to jets and 35% to electron and muons. With the help of the Narrow Width Approximation (NWA) we can have an approximate number of expected events in each possible final state.

These final states are not targeted separately but are grouped by their number of charged leptons. These different channels, characterized by the number of hard electrons and muons identified, are exclusive between each other and have different irreducible backgrounds. We list them in Table 2. We show the total Branching Ratios and expected events for each channel assuming a 139 $fb^{-1}$ luminosity $\mathcal{L}$ without taking into account the efficiencies and acceptances of the analyses themselves. We group together electron and muons as l. Even at LO the minimum number of hard light jets is variable because of the hadronic-decaying taus contribution. Notice that, although the number of true b-quarks is 4, in practice the required number of b-tagged jets is lowered to 2 or 3 b-tags because of the b-tagging efficiency.

| Channel | Final states | BR (%) | Expected events | Main backgrounds |
|---|---|---|---|---|
| All hadronic | $N_j \geqslant 8$, $N_\ell = 0$ | 30 | $499^{+89.8}_{-105}$ | Multijet QCD |
| Monolepton | $N_j \geqslant 7$, $N_\ell = 1$ | 42 | $699^{+126}_{-147}$ | $t\bar{t}$ + Heavy Flavour |
| Opposite Sign Dilepton | $N_j \geqslant 6$, $N_\ell = 2$, $Q_{lep} = 0$ | 15 | $250^{+44.9}_{-52.4}$ | $t\bar{t}$ + Heavy Flavour |
| Same Sign Dilepton | $N_j \geqslant 6$, $N_\ell = 2$, $Q_{lep} = \pm 2$ | 7 | $116^{+30.0}_{-24.5}$ | $t\bar{t}W^{\pm}$, $t\bar{t}H$, $t\bar{t}Z$ |
| Multilepton | $N_j \geqslant 5(4)$, $N_\ell = 3(4)$ | 6 | $99.8^{+18.0}_{-21.0}$ | $t\bar{t}W^{\pm}$, $t\bar{t}H$, $t\bar{t}Z$ |

Table 2: The different 4-top channels characterized by the number of hard charged leptons $\ell = e, \mu$ in the final state. We list the Branching Ratios and the expected events with $\sigma = 11.97^{+18\%}_{-21\%}$, $\mathcal{L} = 139$ $fb^{-1}$ without considering efficiencies and acceptances. We also list the main irreducible backgrounds.

Notice that all channels have missing transverse energy (MET) associated with the hard process, even in the all hadronic channel where the MET comes from the tau neutrino associated with the possible hadronic decaying tau. In practice, the analyses group together monolepton and opposite sign dilepton on one hand same sign dilepton and multilepton on the other to gain statistics because they share irreducible backgrounds. The latter combination presents the best signal-to-background ratio of all possible channels, and thus we focus on it throughout Chapters 3, 4 and 5. Even if the signal-to-background ratio is large, there is still a combination of statistical and systematic uncertainties that the analyses need to overcome. Although the former will reduce considerably after the LHC Run-3, the latter is related to the theoretical uncertainties of the involved hard processes and the instrumental background. As shown in the most recent analyses of ATLAS [25] and CMS [26], the signal modelling with Monte Carlo simulations leads to large systematic uncertainties. The same happens with the modelling of some of the backgrounds. $t\bar{t}W^{\pm}$ in particular shows a persistent

normalization excess. In Chapter 5 we introduce an algorithm that aims to reduce the Monte Carlo dependency of the analyses.

Although the current 4-top measurements are consistent with the SM, the reported measured cross-section central values are still above the expected SM cross-section. Any difference between the SM prediction and the measurements, but specially an excess, could point towards BSM effects. The 4-top cross-section, due to its kinematic structure and the smallness of the SM value, is very sensitive to top-philic light New Physics (NP) that could be produced at the LHC [27, 28]. There exists also in the literature many works that point out 4-top as a sensitive channel to test heavy NP and/or Effective Field Theory effects [29–34]. Because of this, 4-top is a fertile ground for constraining this kind of BSM physics. This is shown in Chapters 2 and 4, where two different models, inspired by different measurements, are both constrained by 4-top physics. The former is a B-Anomalies inspired Leptoquark and the latter is top-philic $Z'$ which seeks to explain the aforementioned $t\bar{t}W^{\pm}$ normalization.

## 1.2 BEYOND THE STANDARD MODEL

Even though the SM is a hugely successful theory, we know it is an incomplete one. There are several phenomena which lie beyond the SM and which should be addressed by considering New Physics (NP), also called Beyond the Standard Model (BSM) physics. Some of these phenomena arise from experimental observations, as is the case of dark matter, neutrino masses and gravity, while others are theoretical in origin as is the case of the $\theta_{QCD}$ problem and the Higgs hierarchy problem. It should be noted however that there is currently no best candidate that can explain these phenomena completely. This is not to say that we should give up on BSM searches but that we should make an effort to be as sensitive as possible to most BSM models. We aim to do this by studying both different BSM models separately and by proposing semi-supervised and unsupervised classification methods to perform model agnostic analyses.

Along this Ph.D. thesis, we are interested in possible BSM imprints at the LHC. Although complete SM extensions are necessary to motivate and guide BSM searches, our main focus is on the possible signatures at the LHC, often trying to be as model agnostic as possible. For this reason, in this Section, we give a brief overview of two specific types of resonances that appear in several BSM theories and which will be the focus of later Chapters without preferring any UV completion in particular. We consider resonances between $\mathcal{O}(100)$ GeV and $\mathcal{O}(1)$ TeV such that resonant production and resonance decay to $t + X$ final states (where $X = t, j, ...$) are relevant.

### 1.2.1 *Leptoquarks*

Leptoquarks are hypothetical scalar or vector fields that can turn quarks into leptons and vice versa [35]. This defining characteristic sets them apart from other proposed BSM fields and yields very interesting and distinct signatures over a wide array of observables, ranging from low-energy flavour physics (e.g. Section 3 of Ref. [35]) to high-energy collider physics where they can be produced resonantly (e.g. Section 5 of Ref. [35]). They are also theoretically well-motivated, arising in several popular extensions of the SM which seek to address some of the problems stated above. Examples of these are Unification scenarios (e.g. Subsection 2.1 of Ref. [35]), neutrino mass

generation (e.g. Subsection 2.2 of Ref. [35]) or to explain the SM flavour structure (e.g. Subsection 2.3 of Ref. [35]).

As they need to couple quarks and leptons, the possible Leptoquarks can be characterized by how they transform under $SU(3)_C - SU(2)_L - U(1)_Y$ in order to keep $q\ell LQ$ invariant under each of the gauge groups transformations. As the leptons are colour singlets, Leptoquarks must transform either in the fundamental 3 or anti-fundamental representation $\bar{3}$ depending on whether they interact with anti-quarks or quarks. Each possible Leptoquark field possesses a unique phenomenology. We focus on two scalar Leptoquark fields, $(\bar{3}, 1, -2/3)$ which we call $S_1$ and $(\bar{3}, 3, -2/3)$ which we call $S_3$. As with the SM fields, we refer to the associated mass eigenstate particles by the name of the field. $S_1$ is a weak singlet and a $Q = 1/3$ particle while $S_3$ is a weak triplet, containing three charged particles $S_3^{4/3}$, $S_3^{1/3}$ and $S_3^{-2/3}$. Defining the charge conjugation operator $C = i\gamma^2\gamma^0$ and $\overline{q}^C = q^T C$, we can write the quark-lepton interaction Lagrangian of $S_1$ and $S_3$ as [35]

$$\mathcal{L}_{S_1} = y_{ij}^{LL} \bar{Q}_L^{C\,i,a} S_1\, \epsilon^{ab}\, \ell_L^{j,b} + y_{ij}^{RR} \bar{u}_R^{C\,i} S_1\, e_R^j + y_{ij}^{\overline{RR}} \bar{d}_R^{C\,i} S_1\, \nu_R^j + \text{h.c.} \tag{1}$$

where $\epsilon$ is the anti-symmetric tensor and $y_{ij}^{LL,RR,\overline{RR}}$ are the $S_1$ couplings constants for left-handed, up-type right-handed and down-type right-handed quark-lepton currents respectively; and

$$\mathcal{L}_{S_3} = y_{ij}\ \bar{Q}_L^{C\,i,a} \epsilon^{ab} (\tau \cdot S_3)^{bc} \ell_L^{j,c} \ + \ \text{h.c.} \tag{2}$$

where $\tau$ is the isovector of Pauli matrices and $y_{ij}$ are the $S_3$ couplings constants for left-handed quark-lepton currents.

Leptoquark phenomenology has experienced a resurgence due to their role as a possible explanation of the B-anomalies. This is detailed briefly before focusing on their collider phenomenology. We only consider scenarios which are relevant to the Leptoquarks couplings and masses studied in Chapters 2 and 3.

*Leptoquarks at low-energy*

Flavour physics is a very active area of research, pushing our understanding of Nature and testing the SM predictions. This is done through low-energy experiments such as LHCb and Belle which measure baryons and mesons observables. These baryons and mesons are composite states of flavour quarks and these experiments have set some of the stringent constraints on the flavour structure of the SM. Due to the non-perturbativity of QCD at low energies, obtaining theoretical predictions for the flavour observables is a challenging task. The most common approach is to use an effective theory described by an effective weak Hamiltonian [36]

$$\mathcal{H}_{\text{eff}} = \sum_i C_i(\mu) \mathcal{Q}_i(\mu) \tag{3}$$

where $\mathcal{Q}_i$ are field operators composed by the elementary particles relevant to energy scale $\mu$ of the observable and $C_i(\mu)$ are the Wilson coefficients of each operator. In this effective theory, the degrees of freedom of the full theory which lie above the energy scale have been integrated out. For the SM, this always includes heavy bosons and the top quark and, depending on the energy scale, could include the bottom and the charm quarks. Although TeV scale Leptoquarks cannot be produced resonantly at

these scales, there are non-resonant effects which can be measured. These include tree-level contributions to observables involving both quarks and leptons such as leptonic and semi-leptonic meson decays, semi-leptonic $\tau$ decays and $\mu - e$ conversion in nuclei; and loop-induced contributions to observables such as electric and magnetic dipole moments of quarks and leptons, rare radiative decays of mesons and leptons, and neutral meson anti-meson oscillations. In our case, we are also integrating $S_1$ or the three components of $S_3$. We approximate the three components of $S_3$ to have the same mass as done in Ref. [35]. Schematically, for a scalar Leptoquark S

$$\mathcal{L}_{\text{eff}}(SM) = \int \mathcal{D}S\mathcal{L}(SM; S) \tag{4}$$

Because to QCD asymptotic freedom, one is able to factorize long distance and short distance behavior. The operators $\mathcal{Q}_i$ contain all long distance behavior while the Wilson coefficientes $C_i$ contain short distance behavior and thus all the information about the integrated degrees of freedom of the full theory. To obtain these coefficients one needs to perform a *matching* of the full and effective theories, which amount to calculate the expectation. This matching will have a theoretical accuracy determined by the perturbation order up to which one can calculate these amplitudes. As one needs to match amputated Green functions, this matching can be done with arbitrary momenta for incoming and outgoing particles. Due to the presence of large logarithms $\alpha_S \text{Log}(M/\mu)$, these matching is usually performed near the scale of the integrated out degrees of freedom $\mu \sim M$. After performing the matching at this scale, the coefficients at a different scale can be computed with the help of the Renormalization Group Equations (RGE) that take into account these large logarithms up until $\mu \sim 1$ GeV. Having obtained the Wilson coefficients for the set of operators, we can use these to obtain theoretical predictions for low-energy processes. All QCD non-perturbative effects are thus confined to the matrix elements of the operators, which need to be calculated by other means [36].

Recently, there have been discrepancies in low-energy measurements in B-factories (that is low-energy experiments focusing on hadrons with at least one bottom quark) which are collectively known as the B-Anomalies. These measurements consist of ratios of meson decays

$$\begin{aligned} R_{K^{(*)}} &= \frac{\mathcal{B}(B \to K^{(*)}\mu\mu)}{\mathcal{B}(B \to K^{(*)}ee)} \\ R_{D^{(*)}} &= \left.\frac{\mathcal{B}(B \to D^{(*)}\tau\bar{\nu})}{\mathcal{B}(B \to D^{(*)}\ell\ell)}\right|_{\ell=e,\mu} \end{aligned}$$

At the time of publication of the work detailed in Chapters 2 and 3, these measurements presented deviations from SM predictions at the $\sim 3.5 - 4\sigma$ level [37–40] for $R_{K^{(*)}}$ and at the $\sim 4\sigma$ level [41–44] for $R_{D^{(*)}}$. The unique flavour structure of Leptoquarks generates tree-level processes that are loop-suppressed at the SM. Because of this, Leptoquarks are a possible explanation to the B-Anomalies [39, 45–51]. If we restrict ourselves to a single Leptoquark scenario, $S_3$ could be responsible for the $R_{K^{(*)}}$ anomaly at tree-level processes [52] while $S_1$ can accommodate both anomalies, accommodating $R_{D^{(*)}}$ at the tree-level and $R_{K^{(*)}}$ at the loop-level [53]. Moreover, one could think of combined explanations which have greater freedom to address these anomalies [51].

It should be noted however than new measurements have reduced the discrepancies. The deviations have reduced to the $\sim 3.1\sigma$ level [54] for $R_{K^{(*)}}$ and to the $\sim 3.1\sigma$ level [55–

57] for $R_{D^{(*)}}$. Even when taking this into account, the B-anomalies are still one of the most interesting deviations from SM predictions as they would imply signatures of possible TeV scale BSM. This BSM could thus be produced at colliders.

*Leptoquarks at Colliders*

Leptoquarks can be produced at colliders through their interactions with SM particles. There are many production mechanisms that could take place in these colliders and which are distinguished by how many on-shell Leptoquarks are produced. We understand as pair production when there are two Leptoquarks on-shell, single-resonant production when there is only one Leptoquark on-shell, and non-resonant all other cases, as for instance a Leptoquark in t-channel.

At the LHC, Leptoquarks are pair produced through their QCD coupling to gluons, as in Fig. 2a). Due to this, pair production is model independent and its usually assumed to be the dominant mode of production. In contrast, single- and non-resonant production are of order $\mathcal{O}(|y|^2)$ and $\mathcal{O}(|y|^4)$, as depicted in Figs. 2b) and 2c) and thus are model dependent.

The Leptoquark is searched through its decay to SM particles. This decay to a quark-lepton pair $ql$ (where $l$ could be either a charged lepton $\ell$ or a neutrino $\nu$) can be calculated to be

$$\begin{aligned}\Gamma_{S\to ql} &= \frac{|y^L_{ql}|^2+|y^R_{ql}|^2}{16\pi M_S^3}(M_S^2-m_q^2-m_l^2)\\ &\quad \sqrt{M_S^4-2M_S^2(m_q^2+m_l^2)+(m_q^2-m_l^2)^2}\end{aligned} \tag{5}$$

where $|y^L_{ql}$ ($y^R_{ql}$) are the left-handed (right-handed) couplings between the scalar Leptoquark and the quark-lepton pair. An useful parametrization is to consider the Branching Ratio of the Leptoquark to $ql$, defined as

$$BR(S\to ql)=\frac{\Gamma_{S\to ql}}{\Gamma_S} \tag{6}$$

where $\Gamma_S$ is the total width of the scalar Leptoquark. This is useful because, if we assume the width is small enough for the NWA to be valid, the resonant Leptoquark production and decay cross-section can be factorized. For pair production

$$\sigma(pp\to S\overline{S}\to q_1l_1\overline{q}_2\overline{l}_2)\approx\sigma(gg\to S\overline{S})BR(S\to\overline{q}_1\overline{l}_1)BR(S\to\overline{q}_2\overline{l}_2) \tag{7}$$

and for single-resonant production

$$\begin{aligned}\sigma(pp\to Sl_2\to\bar{q}_1\bar{l}_1l_2) &\approx \sigma(pp\to Sl_2)BR(S\to\bar{q}_1\bar{l}_1)\\ \sigma(pp\to\bar{S}\bar{l}_2\to q_1l_1\bar{l}_2) &\approx \sigma(pp\to\bar{S}\bar{l}_2)BR(S\to\bar{q}_1\bar{l}_1)\end{aligned} \tag{8}$$

This factorization is not present in non-resonant production.

Leptoquark searches are usually performed by considering each possible quark and charged lepton generation separately. Neutrino flavour is summed out as it is unobservable at the LHC. For example, third-generation $S_1$ pair production searches target the $t\bar{t}\tau^-\tau^+$ and $b\bar{b}+$MET final states exclusively. The reported limits on the observed cross-section are limits on $\sigma BR_1BR_2$ and they are translated to limits on the Leptoquark mass by assuming Branching Ratios of 1 or 0.5, depending on whether the Leptoquark couples to one or two different quark-lepton pairs.

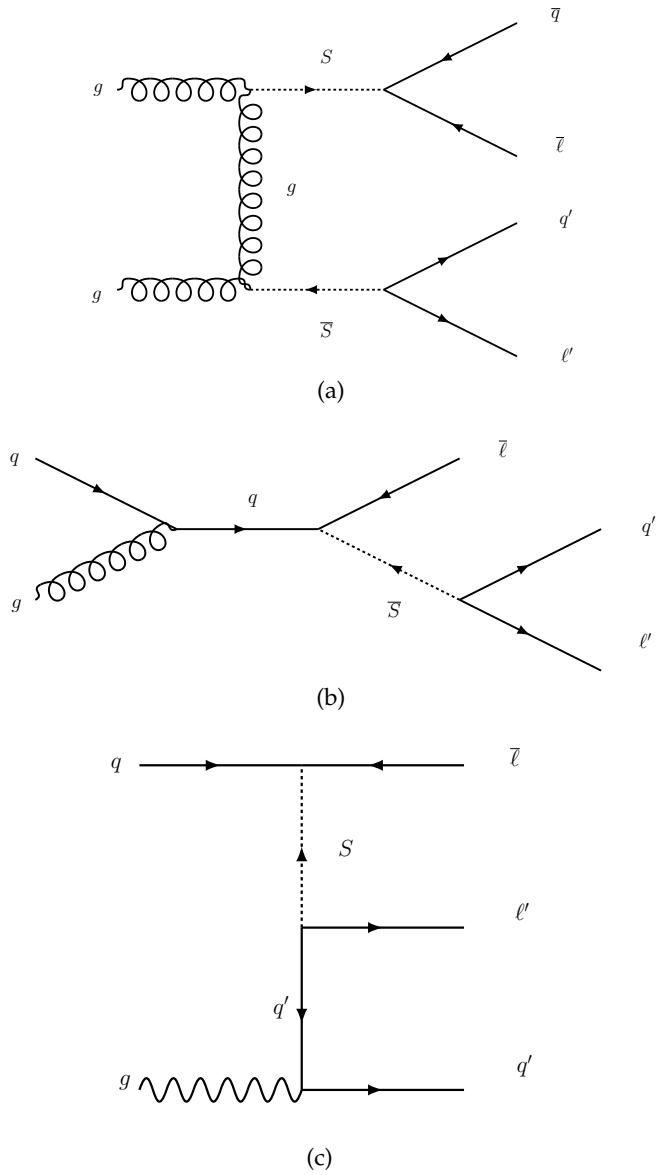



Figure 2: Representative Leptoquark Feynman diagrams for a) pair production b) single-resonant production and c) non-resonant production.

Pair production has been the main target of LHC searches. It is both the most model-agnostic production channel and, under the assumption that $|y_{ij}| << g_s$, the dominant production mode. Chapter 2 is an example of the phenomenology of a Leptoquark model that is pair produced at the LHC. However, as the pair production searches set stringent limits on the Leptoquark mass, single- and non-resonant production have received renewed attention [58–60]. Chapter 3 shows an example where single- and non-resonant production effects are relevant.

### 1.2.2 *Z′ models*

The hypothetical $Z'$ particle is a neutral, colour-singlet spin 1 vector boson. A wide variety of proposed SM extensions have one or more $Z'$ among the BSM resonances and thus it has been consistently and extensively searched for.

Because of its quantum numbers, $Z'$ could in principle mix with the SM Z boson. This mixing is very constrained by electroweak precision tests, e.g. the restrictive bounds imposed by LEP [61]. Because of this, and with an eye towards studying a top-philic $Z'$ in Chapter 4, we neglect all mass mixings. We also neglect couplings to leptons and to *down* type quarks as they are not relevant to explaining the experimental results we consider in Chapter 4.

Its most general interaction Lagrangian with couplings restricted to *up* type quarks is [5]:

$$\mathcal{L}_{int} \supset Z'_\mu \left( \sum_{i,j=u,t,c} \bar{u}_i \gamma^\mu (g^L_{ij} P_L + g^R_{ij} P_R) u_j \right). \tag{9}$$

This Lagrangian does not need to be diagonal in flavour space. Indeed, the presence of flavour changing neutral currents (FCNC) yields a distinct and interesting phenomenology. Although most FCNC are tightly constrained [62, 63], this is not necessarily the case for those involving a top quark. Assuming right-handed only couplings, we study how these FCNC yield a key handle to explain experimental results in Chapter 4. There, we consider several resonant and non-resonant production modes at the LHC. These production modes are fairly specifical to our model due to the chiral nature of the couplings and the fact that we only consider interactions with up-type quarks. Namely, we consider:

1. Resonant $Z'$ production. This is proportional to $|g^R_{uu}|^2$, $|g^R_{cc}|^2$ or $|g^R_{uc}|^2$. The different decay channels of the $Z'$ are constrained by di-jet, $t\bar{t}$ and $tj + \bar{t}j$ resonance searches.

2. Non-resonant $t\bar{t}$ and $tt$ production, proportional to $|g^R_{ut}|^4$, $|g^R_{ut} g^R_{ct}|^2$ or $|g^R_{ct}|^4$. The difference between $t\bar{t}$ and $tt$ productions is the charge of the initial state quarks.

3. $tZ' + \bar{t}Z'$ production. This is proportional to $|g^R_{ut}|^2$ or $|g^R_{ct}|^2$. It is one of the two main channels we consider in Chapter 4 to explain $t\bar{t}W^\pm$ and 4top data.

4. $t\bar{t}Z'$ production. This is proportional to $|g^R_{tt}|^2$. It is one of the two main channels we consider in Chapter 4 to explain $t\bar{t}W^\pm$ and 4top data.

5. $tZ'j + \bar{t}Z'j$ production, where the $Z'$ is emitted from the $t/\bar{t}$ or the u- or c-jet. This channel is suppressed because we consider only right-handed couplings and thus we require a left-handed u to interact with a $W^\pm$ and then become right-handed to interact with the $Z'$. This chirality flip is suppressed by the low up quark mass.

When considering resonant production, the assumed small width of the resonance ensures that the production and decay of the $Z'$ factorize

$$\sigma(pp \to Z' + X \to Y + X) \approx \sigma(pp \to Z' + X) BR(Z' \to Y) \tag{10}$$

In order to have a complete theoretical prediction we need the partial widths, and thus the Branching Ratios, of the $Z'$ to two quarks. When $M > m_i + m_j$, the partial width is

$$\Gamma_{Z'\to q_i\bar{q}_j} = \frac{|g^R_{ij}|^2}{16\pi M^3_{Z'}}\sqrt{M^4_{Z'} - 2M^2_{Z'}(m^2_{q_i}+m^2_{q_j}) + (m^2_{q_i}-m^2_{q_j})^2}$$
$$(2M^2_{Z'}(1 - \frac{(m^2_{q_i}-m^2_{q_j})^2}{2M^4_{Z'}}) - m^2_{q_i} - m^2_{q_j}) \quad (11)$$

that reduces to $\frac{N_c}{24\pi}M_{Z'}|g^R_{ij}|^2$ for massless quarks ($N_c = 3$ is the number of quark colours). For flavour changing neutral currents where $i \neq j$, there are two partial widths: $\Gamma_{Z'\to q_i\bar{q}_j}$ and $\Gamma_{Z'\to q_j\bar{q}_i}$.

If $M_{Z'} < m_i + m_j$, then at least one of the quarks must be produced off-shell. For our purposes, we are interested in the $Z' \to t\bar{t}$ decay when $m_t < M_{Z'} < 2m_t$. In this case, one of the tops is produced off-shell and decays to a $W^\pm$ boson and a b-quark, with appropriate charges, resulting in a three-body decay. If the mass is low enough, one should consider the case where both top quarks are produced off-shell. This $WbWb$ decay width can be calculated as [64]

$$\Gamma_{Z'\to W^+bW^-\bar{b}} = \frac{|g^R_{tt}|^2}{4\pi M^5_{Z'}}\frac{\alpha^2_{EW}}{(32\pi)^2}\frac{m^4_t}{\sin^4(\theta_W)m^4_W}$$
$$\int_{(m_W+m_b)^2}^{(M_{Z'}-m_W-m_b)^2}\frac{dQ^2_1}{(Q^2_1)^2}\int_{(m_W+m_b)^2}^{(M_{Z'}-\sqrt{Q^2_1})^2}\frac{dQ^2_2}{(Q^2_2)^2}$$
$$F(M^2_{Z'},Q^2_1,Q^2_2)\frac{F(m_W,Q^2_1,m^2_b)}{(Q^2_1-m^2_t)^2+m^2_t\Gamma^2_t}\frac{F(m^2_W,Q^2_2,m^2_b)}{(Q^2_2-m^2_t)^2+m^2_t\Gamma^2_t} \quad (12)$$

where $F(a,b,c) = \sqrt{a^2-2a(b+c)+(a-b)^2}(a(2a-(b+c))-(b-c)^2)$ is an auxiliary function, $\alpha_{EW}$ is the fine structure constant, $\theta_W$ is the Weinberg angle and $\Gamma_t = -\frac{\alpha_{EW}}{16}\frac{F(m^2_W,m^2_t,m^2_b)}{\sin^2(\theta_W)m^2_W m^3_t}$ is the top-quark width.

The $tWb$ decay width can be obtained by setting one top quark on-shell, which is done by taking the NWA for one of the tops $Q^2$ distribution:

$$\frac{1}{(Q^2_1-m^2_t)^2+m^2_t\Gamma^2_t} \approx \frac{\pi}{m_t\Gamma_t}\delta(Q^2_1-m^2_t) \quad (13)$$

with which we get

$$\Gamma_{Z'\to tW^-\bar{b}} = \frac{|g^R_{tt}|^2}{4\pi M^5_{Z'}}\frac{\alpha^2_{EW}}{(32\pi)^2}\frac{1}{\sin^4(\theta_W)m^4_W}$$
$$\frac{\pi}{m_t\Gamma_t}\int_{(m_W+m_b)^2}^{(M_{Z'}-m_t)^2}\frac{dQ^2_2}{(Q^2_2)^2}$$
$$F(M^2_{Z'},m^2_t,Q^2_2)F(m^2_W,m^2_t,m^2_b)\frac{F(m^2_W,Q^2_2,m^2_b)}{(Q^2_2-m^2_t)^2+m^2_t\Gamma^2_t} \quad (14)$$

The reader can verify that putting the remaining top quark on-shell (assuming $M_{Z'} > 2m_t$) yields the expected decay width. In practice, we compute the three-body (and if necessary four-body) decay width numerically with `MadWidth`, a `MadGraph` module [65].

The $Z'$ can also be constrained indirectly. For example, the $Z'$ would be responsible for rare top quark decays. These rare decays have been studied in Ref. [66], obtaining:

$$BR(t\to u_jV) = C_F\frac{m^3_t}{8\pi\Gamma_t}\left(1-\frac{m^2_j}{m^2_t}\right)(|F^{tj}_M|^2+F^{tj}_E\|^2) \quad (15)$$

where $C_F = 1\ (4/3)$ when V is photon $\gamma$ (gluon g), and $F_M^{tj}$ and $F_E^{tj}$ are auxiliary functions define in Ref. [66] that depend on the coupling structure and the $Z'$ mass. These decays can be used to constrain the coupling structure and $Z'$ mass with the help of the experimental bounds set in Refs. [67–70].

At lower energies, the $Z'$ is constrained by a wide array of experimental results. In Chapter 4 we calculate the one relevant flavour constrain to our model, $D^0 - \bar{D}^0$-meson mixing.

## 1.3 TOPIC MODELING TECHNIQUES FOR MACHINE LEARNING AT THE LHC

As we mentioned in Section 1.1.1, in this Thesis we propose to use two topic modeling techniques in LHC searches. Topic modeling is a subject of natural language processing and Machine Learning that concerns the study of statistical models to recover abstract topics that occur in a corpus of documents.

Although Machine Learning techniques are well established as useful tools in LHC physics, the main focus has been to apply discriminative algorithms to optimize data selection for supervised searches. The two main algorithms in the LHC experimental toolkit are Boosted Decision Trees and Neural Networks (NN) [71] with both relying on (an usually large set of) Monte Carlo simulations to improve upon cut-based techniques. However, applications of semi-supervised and unsupervised techniques to SM analyses and BSM searches have received increased attention in the last few years, see e.g. refs. [72–98]. These techniques are designed to separate signal and background events in mixed samples, and could therefore be run directly on experimental data without the need for pure Monte Carlo training samples.

Our main goal is to build upon the existing topic modeling literature to perform semi-supervised and/or unsupervised classification of the LHC data. In doing this, we hope to reduce the theoretical bias and Monte Carlo uncertainties that difficult current analyses. We also are focused on obtaining an unsupervised algorithm that allows to search for a potentially vast array of BSM signals. This is necessary because, as detailed in Subsection 1.2, there is no current preferred BSM model to guide our searches.

As its name suggests, a topic model is an statistical model where we assume a collection of documents, which we call the corpus, can be reduced in some way to a set of abstract topics. These abstract topics then encode the corpus information in a lower dimensional latent space which can be exploited to characterize the corpus by any means possible, e.g. clustering documents according to their topic composition.

For example, Latent Semantic Analysis [99] (LSA) is a topic model where we compute the word occurrence per document matrix X and using truncated Singular Value Decomposition of dimension k we obtain the best approximation (that has the lowest error using the Frobenius norm) of X using k parameters.

However, topic models like LSA are not appropriate to study LHC physics. Instead of abstract topics, we are looking for meaningful topics that have a clear relation to the hard processes we wish to study. Because of this we consider generative probabilistic topic models. Generative models assume the existence of latent variables from which the observed data is sampled and from which we could sample "new" data. The specific model with its conditional structure is best encoded in a graphical model [71]. Instead of point estimates of the topics, we are interested in the posterior probabilities of the topics given the observed corpus. This can be obtained by numerical methods such as Gibbs sampling or Variational Inference. Another benefit of probabilistic

models is that, as we consider the full posterior, the overfitting reduces significantly. A more detailed description of probabilistic topic models and the variational bayes inference method used to obtain the topics can be found in Chapter 6.

We consider two probabilistic topic models that have already been adapted to the LHC. We consider how the *Topic Model Demixer algorithm* or *Demixer algorithm* introduced in Ref. [82] can be applied to 4-top physics in Chapter 5. We expand on the use of Latent Dirichlet Allocation or LDA introduced in Ref. [100], performing a thorough exploration of the hyperparameters of the model to define selection criteria needed for a full-on unsupervised search in Chapter 6. Both of them need to adapt the topic models for the LHC by defining documents and words. For example, in Chapter 5 documents are events and each of them consists of only one word. In Chapter 6 the documents are again the events but now the words are a set of jet substructure observables, with each subjet splitting an observed word.

# Part I

# DEDICATED MONTE CARLO-BASED SEARCHES

# 2

# LEPTOQUARK PAIR PRODUCTION AT THE LHC

In this Chapter we explore the LHC phenomenology of a Leptoquark that could be responsible for the $R_K^{(*)}$ anomaly. As shown in Ref. [39], the tree-level exchange of scalar or vector Leptoquarks at the TeV scale generate dimension-six operators with Wilson coefficients of appropriate size to accommodate the deviations observed in LFU observables in B-meson decays. A very interesting possibility is to consider a new strongly coupled field theory (SCFT) that, besides the Higgs boson, generates resonances at the TeV scale, some of which are Leptoquarks with the characteristics described in Section 1.2.1.

We show below how, in the context of anarchic partial compositeness, $R_{K^{(*)}}$ measurements prefer a Leptoquark mass of $\lesssim$ TeV. However, electroweak precision tests (EWPT) in general require the masses of resonances to be $m_* \gtrsim$ few TeV, thus a separation between the lightest Leptoquark mass and $m_*$ is required. This splitting can be obtained if the Leptoquark involved in $R_{K^*}$ is a pseudo Nambu-Goldstone boson (pNGB), since in this case its mass is given by $m_{pNGB} \simeq \frac{g_{SM}}{4\pi} m_*$, where $g_{SM}$ is an SM coupling such as the gauge couplings or the top quark Yukawa coupling. Therefore, in the following we propose that this Leptoquark is a pNGB of a new sector, and we study the case of $S_3$.

Considering the Leptoquark triplet as a pNGB allows a splitting between the mass of $S_3$ and the masses of other NP states, which can then be made naturally heavier and whose effects in EWPT can be neglected. A suitable framework for this scenario is a new strongly coupled sector with resonances heavier than a few TeV [46, 52, 101], and with a global symmetry spontaneously broken by the strong dynamics, generating a set of pNGB's. Within the pNGB's there is the state $S_3$, as well as a composite Higgs boson. There might exist also additional light scalars, but we assume their phenomenological impact in the context of this study can be neglected. We also assume anarchic partial compositeness, which leads to a well defined pattern of couplings, which we take as a benchmark point.

We focus on the component of $S_3$ with electric charge 1/3, denoted $S_3^{1/3}$, and its LHC phenomenology. The main decay modes are $S_3^{+1/3} \to \bar{t}\,\ell^+$ and $\bar{b}\bar{\nu}$, with a preference for second and third generation leptons. We consider several recent LHC searches that are sensitive to the presence of the $S_3^{1/3}$ state and derive constraints in its Branching Ratios as a function of mass.

This Chapter is organized as follows. In Section 2.1 we present an effective description that allows to model the Leptoquarks as composite states and we discuss the phenomenology of the model, combining model predictions with the available B-physics anomalies results. In Section 2.2 we discuss and implement the re-casting of three LHC searches to constrain the Leptoquark's parameter space. Finally, in Section 2.3 we provide a summary and outlook.

The basis of this Chapter is the collaboration with Ezequiel Alvarez, Leandro Da Rold, Aurelio Juste and Tamara Vazques Schroeder that resulted in Ref. [102]. I participated in the recasting of the searches to limits in the BRs space as reported in

Section 2.2, particularly in performing and tuning the simulations needed for 4-top recasting, with the exception of the likelihood fit performed for the 4-top simulations.

## 2.1 A MODEL FOR A B-ANOMALIES INSPIRED PNGB LQ

We present an effective description that allows to model the Leptoquarks as composite states in Subsection 2.1.1. We discuss the phenomenology of the model in Subsection 2.1.2, where we combine the model predictions with the available B-physics anomalies results and we detail its phenomenology and the impact of different observables on the coupling structure.

### 2.1.1 *Leptoquarks as a pNGB from an SCFT*

We consider an SCFT that produces resonances with masses $m_* \gtrsim$ few TeV of the kind detailed in Ref. [103]. The SCFT has a global symmetry G, spontaneously broken by the strong dynamics to a subgroup H containing the SM gauge group. The Higgs boson and the $S_3$ Leptoquark are pNGBs associated to this spontaneous breaking of global symmetries. This assumption determines the size of the couplings between the SM fermions and the SCFT. There might be other pNGB states in the coset G/H, Leptoquarks or not, but we will assume that their effect is subdominant for the phenomenology that we will study, compared with the effect of $S_3$. The flavour structure of the SCFT is taken to be anarchic [104, 105], meaning that there are no preferred directions in flavour space, and thus all the coefficients of tensors in flavour space are of the same order.

In the model we consider, the SM states are elementary with the SM gauge fields weakly gauging a subgroup of the global symmetries of the SCFT. The SM fermions have linear interactions with the SCFT:

$$\mathcal{L} \supset \omega\, \bar{\psi}_{SM} \mathcal{O}_{SCFT}$$

where $\omega$ is the coupling at the UV scale at which this interaction is defined. Bilinear coupling effects are taken to be subdominant compared with the effect of the linear ones, and thus not considered in the following. All these interactions explicitly break the global symmetry of the SCFT, resulting in a potential being induced at loop level for the resulting pNGBs. The top quark dominates the contributions to the potential, and it can trigger the breaking of the electroweak symmetry.

Rather than attempting to develop a fundamental theory, we consider an effective theory describing the resonances and their interactions. For simplicity we adopt the one scale and one coupling description of the resonances [106], defining $g_*$ as the typical size of the coupling between the resonances. We assume $g_*$ to be larger than the SM couplings but still in the perturbative regime, i.e $1 < g_* < 5$. At low energies the linear interactions of the SM fermions lead to mixing with the resonances of the SCFT:

$$\mathcal{L}_{eff} \supset \lambda\, f\, \bar{\psi}_{SM} \Psi_{SCFT}\,, \tag{16}$$

where $\lambda$ is the coupling at the scale $m_*$, $f = m_*/g_*$ is the pNGB decay constant, and $\Psi_{SCFT}$ is a fermionic vector-like resonance of the composite sector. This mechanism is usually known as partial compositeness [107], since the mass eigenstates are a mixture

of elementary and composite states, with degree of compositeness $\epsilon \sim \lambda/g_*$. Fermionic resonances and the Higgs have Yukawa interactions:

$$\mathcal{L}_{\text{eff}} \supset g_* c_{ij} \bar{\Psi}_{\text{SCFT}} H \Psi'_{\text{SCFT}},$$

with $c_{ij}$ anarchic coefficients in flavour space of $\mathcal{O}(1)$.

In the case of an anarchic SCFT, linear interactions lead to what is usually referred to as "anarchic partial compositeness" (APC). Assuming that the scale evolution of the linear couplings is driven by the dimension of the SCFT operator and that there exists a separation between the UV scale of Eq. 16 and $m_*$, it is possible to obtain a mixing hierarchy of the SM fermions. After electroweak symmetry breaking, this mixing leads to interactions with the Higgs boson that generate masses for the SM fermions. Their Yukawa couplings have the structure:

$$y_{ij} \simeq g_* \epsilon_{Li} \epsilon_{Rj} c_{ij},$$

where $i, j$ are generation indices and $\epsilon_{Li,Ri}$ is the degree of compositeness of the Left-handed (LH) or Right-handed (RH) quirality. Large masses, as in the case of the top quark, can be obtained by taking the left-handed (LH) and right-handed (RH) mixing to be of $\mathcal{O}(1)$. This in turn implies a large degree of compositeness for the top quark, $\epsilon \sim \mathcal{O}(1)$. In contrast, tiny masses can be obtained by taking the mixing of one of the chiralities to be small, as in the case of the light quarks and leptons. A hierarchy of mixing of the LH quarks can also lead to the CKM matrix. The scenario of APC also provides a very economic mechanism to satisfy most of the flavour bounds present in Composite Higgs Models [105, 108]. Of these constraints, the most important ones arise from the kaon system and electromagnetic dipole moments, although some solutions have been proposed [109–111].

Furthermore, the masses and weak mixing angles of the SM quarks can be reproduced by taking:

$$\begin{aligned}
&\epsilon_{q_1} \sim \lambda_C^3 \epsilon_{q_3}\,, \qquad \epsilon_{u_3} \sim \frac{y_t^{SM}}{g_* \epsilon_{q_3}}\,, \qquad \epsilon_{u_2} \sim \frac{y_c^{SM}}{\lambda_C^2 g_* \epsilon_{q_3}}\,, \qquad \epsilon_{u_1} \sim \frac{y_u^{SM}}{\lambda_C^3 g_* \epsilon_{q_3}}\,, \\
&\epsilon_{q_2} \sim \lambda_C^2 \epsilon_{q_3}\,, \qquad \epsilon_{d_3} \sim \frac{y_b^{SM}}{g_* \epsilon_{q_3}}\,, \qquad \epsilon_{d_2} \sim \frac{y_s^{SM}}{\lambda_C^2 g_* \epsilon_{q_3}}\,, \qquad \epsilon_{d_1} \sim \frac{y_d^{SM}}{\lambda_C^3 g_* \epsilon_{q_3}}\,,
\end{aligned} \tag{17}$$

where $\epsilon_{q_i}$ corresponds to the LH doublet of $i$th generation, whereas $\epsilon_{u_i}$ and $\epsilon_{d_i}$ to the RH singlets, $\lambda_C$ is the Cabibbo angle, and $y^{SM}$ are the Yukawa couplings of the SM. The degree of compositeness of all the quarks is determined by physical quantities up to a common factor $1/(g_* \epsilon_{q_3})$. The first column leads to the CKM matrix

$$V_{CKM} \approx \begin{pmatrix} 1 - \frac{\lambda_C}{2} & \lambda_C & \lambda_C^3 \\ \lambda_C & 1 - \frac{\lambda_C}{2} & \lambda_C^2 \\ \lambda_C^3 & \lambda_C^2 & 1 \end{pmatrix}$$

while the other three rows lead to the quark masses, as detailed in Chapter 4 of Ref. [112].

The lepton sector depends on the realization of neutrino masses. We will assume that the angles of the PMNS matrix are generated by the matrix that diagonalizes the neutrino mass matrix. For the charged leptons we will take hierarchical mixing of the

same size for both chiralities of each generation. This choice minimizes the constraints from flavour violating transitions in the lepton sector [112]. Therefore, we assume:

$$\epsilon_{\ell_j} \sim \epsilon_{e_j} \sim \sqrt{\frac{m_{e_j}}{g_* v}}\,, \qquad j = 1,2,3\,, \tag{18}$$

where $v$ stands for the vacuum expectation value (vev) of the Higgs field.

The interactions between the SM fermions and the resonances depend explicitly on the mixing. Although the SCFT is anarchic, the couplings with the resonances are not because they are mediated by the hierarchical structure of the mixing. Interactions involving two SM fermions, $\psi$ and $\psi'$, as well as a spin 0 or 1 resonance, are expected to be of order $\epsilon_\psi g_* \epsilon'_\psi$. Roughly speaking, heavy SM fermions, mainly the top quark but also the left-handed bottom quark and eventually the $\tau$ lepton, will have sizable couplings whereas the coupling of the light SM fermions to the resonances will in general be suppressed.

For the case of a spin 0 $S_3$ Leptoquark resonance, in Section 1.2.1 we wrote its interaction Lagrangian as [35]:

$$\mathcal{L}_{\text{int}} \supset y\ \bar{Q}_L^{C\,a} \epsilon^{ab} (\tau \cdot S_3)^{bc} \ell_L^c \ + \text{h.c.}\,, \tag{19}$$

It is useful to define the electric charge eigenstate basis: $S_3^{+4/3} = (S_3^1 - iS_3^2)/\sqrt{2}$, $S_3^{+1/3} = S_3^3$ and $S_3^{-2/3} = (S_3^1 + iS_3^2)/\sqrt{2}$. Expanding Eq. 19 in components and rotating to the mass basis of fermions we obtain [35]:

$$\begin{aligned}\mathcal{L}_{\text{int}} \supset &-\sqrt{2}\, y_{jk}\ \bar{d}_L^{c\,j}\ S_3^{+4/3}\ e_L^k + \sqrt{2}\ (V^T y U)_{jk}\ \bar{u}_L^{c\,j}\ S_3^{-2/3}\ \nu_L^k \\ &- S_3^{+1/3}\ [(yU)_{jk}\ \bar{d}_L^{c\,j}\ \nu_L^k + (V^T y)_{jk}\ \bar{u}_L^{c\,j}\ e_L^k] \ + \text{h.c.}\,,\end{aligned} \tag{20}$$

where $V$ and $U$ are the CKM and PMNS matrices.

The SM gauge symmetry is compatible with the presence of $S_3 qq$ interactions which in turn can mediate proton decay. The presence of these interactions with Leptoquark masses of order TeV would rule out the present scenario. We assume that there exists an additional symmetry that forbids these $S_3 qq$ interactions [39] while allowing $S_3 q\ell$ interactions. This can be achieved with a discrete $Z_2$ symmetry, e.g by assigning an odd parity ($-$) to $S_3$ and $q$, and an even parity ($+$) to $\ell$.

The $S_3$ Leptoquark also interacts with the Higgs boson. It is interesting to study these interactions because they can split the masses of the components of $S_3$. Considering up to dimension-four operators, there are two independent terms of type $H^2 S_3^2$ [35]. Evaluating the Higgs field on its vev, they induce splitting between $S_3^j$. In the presence of other Leptoquarks, as $S_1$, $\tilde{S}_2 \sim (3,2)_{1/6}$ and $\hat{S}_2 \sim (3,2)_{-5/6}$, new effects are present at the level of dimension-four operators. The splitting between components of a given multiplet is of order $v$, thus $\Delta m_S \sim \mathcal{O}(100\ \text{GeV})$ can be expected.

The couplings $y$ in Eq. 19 can be estimated by considering Eqs. 17 and 18, obtaining

$$y \simeq \epsilon_{q_3} \sqrt{g_*} \begin{pmatrix} \lambda_C^3 \sqrt{m_e/v} & \lambda_C^3 \sqrt{m_\mu/v} & \lambda_C^3 \sqrt{m_\tau/v} \\ \lambda_C^2 \sqrt{m_e/v} & \lambda_C^2 \sqrt{m_\mu/v} & \lambda_C^2 \sqrt{m_\tau/v} \\ \sqrt{m_e/v} & \sqrt{m_\mu/v} & \sqrt{m_\tau/v} \end{pmatrix}\,, \tag{21}$$

where $\epsilon_{q_3}$ is a coupling of $\mathcal{O}(1)$ and $\lambda_C$ is the Cabibbo angle. A numerical estimate of $y$, up to the factor $1/(\epsilon_{q_3}\sqrt{g_*})$, can be found in Ref. [39]. Without the need of numerical estimates one can see that the model presents a clear hierarchy by construction

$$\frac{y_{3i}}{5y_{2i}} \sim \frac{y_{3i}}{25y_{1i}} \sim 1$$
$$\frac{y_{i3}}{4y_{i2}} \sim \frac{y_{i3}}{60y_{i2}} \sim 1$$

for $i = 1, 2, 3$. Motivated by this, we study the $S_3$'s LHC Phenomenology assuming only $y_{32}$, $y_{33}$ couplings, $|V_{33}| = 1$, an unitary PMNS matrix and all fermions massless.

### 2.1.2 $S_3$ *LHC Phenomenology*

Having detailed the $S_3$ interaction Lagrangian and the corresponding estimates of the couplings, we have the tools to study its phenomenology at the LHC. We find that the main decay modes include a third-generation quark and are

$$S_3^{-2/3} \rightarrow \sum_i \bar{t}\bar{\nu}_i \quad \propto \quad |y_{33}|^2 + |y_{32}|^2 \tag{22}$$

$$S_3^{+1/3} \rightarrow \bar{t}\tau^+,\ \bar{t}\mu^+, \sum_i \bar{b}\bar{\nu}_i \quad \propto \quad |y_{33}|^2,\ |y_{32}|^2,\ |y_{33}|^2 + |y_{32}|^2 \tag{23}$$

$$S_3^{+4/3} \rightarrow \bar{b}\tau^+,\ \bar{b}\mu^+ \quad \propto \quad |y_{33}|^2,\ |y_{32}|^2 \tag{24}$$

and their CP conjugates. We have indicated, along with each decay mode, the corresponding couplings involved. The exact formulas including the CKM mixing matrix are deduced from Eq. 20.

Note that, although $S_3^{4/3}$ is the component that would be involved in the neutral-current B-physics anomaly $R_K^{(*)}$, its main decay mode ($S_3^{4/3} \rightarrow b\tau$) is hard to detect at the LHC when produced on-shell because of difficulties in $\tau$-lepton tagging and large associated backgrounds. Recent dedicated searches [113] and constraints from $B - L$ R-parity violating stops [114], are used to set limits on this particle. In contrast, the $S_3^{2/3}$ component of the triplet has a unique decay mode, $S_3^{2/3} \rightarrow t\nu$, which is quite well constrained by stop searches [115]. On the other hand, $S_3^{1/3}$ is the only component that has decay channels to the upper and lower members of the lepton doublet. The $S_3^{1/3} \rightarrow b\nu$ decay mode is constrained by sbottom searches [115], whereas the charged lepton decay mode has limits from dedicated Leptoquark searches [116, 117] and can also be constrained via searches in multilepton plus jets final states, such as 4-top ($t\bar{t}t\bar{t}$) searches [118]. Because of this, in this Chapter we choose to focus on the constraints on the $S_3^{1/3}$ Leptoquark.

It is interesting to notice in Eqs. 22-24 that these decays are driven by what our framework predicts to be the largest $y_{ij}$. In contrast, because of kinematic considerations, in B-meson decays most of these $y_{3j}$ are not accessible unless they are accompanied by a suppression factor $y_{2j}$, as for instance $y_{22}$ in $B \rightarrow K\mu^+\mu^-$.

In addition to the different phenomenology of these decay channels, we should expect a splitting of the mass eigenstate due to the electroweak symmetry breaking. As detailed in Ref. [103], we can expect a Leptoquark mass splitting of $\Delta m_S \sim \mathcal{O}(100\,\text{GeV})$.

As we detailed in Section 1.2.1, the $S_3$ Leptoquark couples to gluons through the strong interaction with coupling strength $g_s$ due to its colour charge and thus can be pair produced at the LHC independently of its weak coupling to quark and leptons.

Moreover, depending on the Leptoquark mass and this weak coupling, there will be some characteristic mass limit above which it would be more promising to study single, rather than pair production [51, 119]. Given the current status of the B-physics anomalies, we find that a $S_3$ with couplings motivated by partial compositeness has a larger cross-section for pair production than for single production for Leptoquark masses $< 1 - 1.5$ TeV (see for instance Ref. [46]). Therefore, in this Chapter we consider solely QCD-mediated Leptoquark pair production.

Once Leptoquarks are pair produced at the LHC, they decay into quarks and leptons. In the NWA, Leptoquark production and decay factorize and the decay is parametrized by the Branching Ratios. These Branching Ratios are determined by the relevant $y_{ij}$ coefficients. The $R_K^{(*)}$ anomalies suggest the following relationship [39]:

$$\frac{y_{32}\, y_{22}^* - y_{31}\, y_{21}^*}{M^2} \approx \frac{1}{(33\ \text{TeV})^2}\ . \tag{25}$$

Since in the context of partial compositeness the couplings involving fermions of the first generation are suppressed compared with those involving the second-generation, Eq. 25 simplifies to

$$\frac{y_{32}\, y_{22}^*}{M^2} \approx \frac{1}{(33\ \text{TeV})^2}\ . \tag{26}$$

Within our framework there are further expected relations between the $y_{ij}$ coefficients, which determine a preferred curve in parameter space. In particular, it is useful to recall the following relationships:

$$\frac{y_{22}}{y_{32}} \sim \lambda_C^2 \approx 0.05\ , \tag{27}$$

$$\frac{y_{33}}{y_{32}} \sim \sqrt{\frac{m_\tau}{m_\mu}} \approx 4\ , \tag{28}$$

which come from Eq. 21.

Within our framework, with no other decay channel other than $S_3^{1/3} \to t\tau, t\mu, b\nu$, and again assuming only $y_{33}$, $y_{32}$ couplings, $|V_{33}| = 1$, an unitary PMNS matrix and all fermions massless, we can obtain the following Branching Ratios:

$$\text{BR}(S_3^{1/3} \to t\tau) = \frac{|y_{33}|^2}{2(|y_{33}|^2 + |y_{32}|^2)}\ , \tag{29}$$

$$\text{BR}(S_3^{1/3} \to t\mu) = \frac{|y_{32}|^2}{2(|y_{33}|^2 + |y_{32}|^2)}\ , \tag{30}$$

$$\text{BR}(S_3^{1/3} \to b\nu) = \frac{1}{2}\ . \tag{31}$$

Interestingly, due to CKM and PMNS unitarity and the chosen $SU(2)_L$ structure of the interaction Lagrangian, the $\text{BR}(S_3^{1/3} \to b\nu)$ is fixed and

$$\text{BR}(S_3^{1/3} \to t\tau) + \text{BR}(S_3^{1/3} \to t\mu) = \text{BR}(S_3^{1/3} \to b\nu) = \frac{1}{2}\ . \tag{32}$$

In a more general model, this relationship may not hold depending on the assumptions made. In models without other decay channels, giving a maximum value for $\text{BR}(S_3^{1/3} \to b\nu)$ can set a minimum value for $\text{BR}(S_3^{1/3} \to t\tau) + \text{BR}(S_3^{1/3} \to t\mu)$. In

our model, it also sets a maximum value to $\mathrm{BR}(S_3^{1/3} \to t\tau) + \mathrm{BR}(S_3^{1/3} \to t\mu)$. By taking Eq. 28 and considering the preferred region to be within 50% of the central value reported there, we take

$$\frac{y_{33}}{y_{22}} = 4 \pm 2,$$

and obtain that for our relevant component $S_3^{1/3}$

$$\frac{\mathrm{BR}(S_3^{1/3} \to t\tau)}{\mathrm{BR}(S_3^{1/3} \to t\mu)} = (4 \pm 2)^2 \,. \tag{33}$$

The size of the preferred region is somewhat arbitrary and it reflects the fact that Eq. 28 is determined up to factors of $\mathcal{O}(1)$.

Eqs. 32 and 33 define a curve in the Branching Ratio parameter space in which our pNGB-based Leptoquark model is preferred. In Section 2.2 we recast several existing searches to derive bounds on the couplings and mass of $S_3^{1/3}$, and compare them to the preferred curve within our model. We also discuss the constraints resulting from searches for the other members of the triplet. However, one should bear in mind the possibility of a significant mass splitting, as indicated above.

## 2.2 RECASTING LHC SEARCHES

In this Section we study different LHC searches that can be recasted to constrain our parameter space. We detail the different searches in Subsection 2.2.1 and present the constraints they set in Subsection 2.2.2.

### 2.2.1 *Searches*

In a partial compositeness framework, as in many other BSM scenarios, the main decay channels of $S_3^{1/3}$ are expected to be $t\tau$ and $b\nu$, with also a non-negligible contribution from $t\mu$. We consider a dedicated search for pair production of $S_3^{1/3}$ decaying into $t\tau$ [116], as well as two recent searches in final states with $b\bar{b} + \mathrm{MET}$ [115], and multilepton plus b-jets (in the context of a 4-top search) [118], which would be sensitive to $S_3^{1/3}$. The CMS Collaboration has also released a direct search for pair production of $S_3^{1/3}$ decaying into $t\mu$ [117], but unfortunately the public document does not contain sufficient information for a reinterpretation, and thus could not be considered in our study. In this Section we pay special attention to the 4-top search, not only because it had not been analyzed in this context yet but also because its translation to the Leptoquark parameter space requires a detailed recast of the experimental results. Possible constraints coming from other members of the Leptoquark triplet are mentioned in Section 2.3. Some details of each of the searches considered are provided below:

- *Leptoquark search in the $t\bar{t}\tau^+\tau^-$ channel:* The CMS Collaboration has recently performed a dedicated Leptoquark search in this channel using 35.9 $\mathrm{fb}^{-1}$ of pp collision data at $\sqrt{s} = 13$ TeV [116]. Events are required to have one light lepton ($\ell = e, \mu$), at least one hadronically decaying tau candidate ($\tau_{\mathrm{had}}$), and at least two jets. Different event categories are considered: $\ell+\tau_{\mathrm{had}}$ with opposite charge and at least four jets, $\ell+\tau_{\mathrm{had}}$ with same charge, and $\ell+\geqslant 2\tau_{\mathrm{had}}$ (with an opposite-charge $\tau_{\mathrm{had}}$ candidate pair). In the categories with exactly one $\tau_{\mathrm{had}}$ candidate,

the kinematic reconstruction of the top-quark candidate is performed. The main discriminating variable is the $p_T$ of the top-quark candidate. In contrast, in the category with $\geqslant 2\tau_{had}$ candidates a counting experiment is performed. Upper limits on $\sigma(pp \to S_3^{+1/3} S_3^{-1/3}) \times BR(S_3^{1/3} \to t\tau)^2$ are obtained and compared to the theoretical prediction as a function of the Leptoquark mass $M_{S_3^{1/3}}$. For $BR(S_3^{1/3} \to t\tau) = 1$, the 95% C.L. lower limit on the $S_3^{1/3}$ mass is $M_{S_3^{1/3}} > 0.9$ TeV. We recast this result as a function of the different Branching Ratios considered. Because we consider Leptoquarks that can also decay to muons and given the experimental requirements summarized above, we make the approximation that the acceptance is similar for signal events with $t\tau t\tau$ and $t\tau t\mu$ final states, whereas negligible otherwise.

- *Sbottom search in the $b\bar{b}$ + MET final state:* For a massless neutralino, the process $pp \to \tilde{b}\bar{\tilde{b}} \to b\bar{b}\tilde{\chi}_1^0\bar{\tilde{\chi}}_1^0$ has the same final state signature as $pp \to S_3^{1/3}\bar{S}_3^{1/3} \to b\bar{\nu}b\nu$. Therefore, searches for direct sbottom production at the LHC [115, 120] are well suited to probe the $S_3^{1/3} \to b\nu$ decay mode. The most restrictive 95% CL lower limit on the bottom mass for a massless neutralino, $m_{\tilde{b}} > 1.18$ TeV, is obtained by the CMS Collaboration using 35.9 $fb^{-1}$ of pp collision data at $\sqrt{s} = 13$ TeV [115]. This limit would be translated to the identical bound for the $S_3^{1/3}$ mass under the hypothetical assumption that $BR(S_3^{1/3} \to b\nu) = 1$, and the value of theory cross section at this mass would then represent the experimental upper limit on the cross section. Considering the Leptoquark model in this Chapter, this Branching Ratio is reduced because of the presence of new decay modes which are assumed to have zero acceptance in this analysis. We estimate the appropriate lower mass limit from the crossing of the new theory prediction as a function of mass with the estimated experimental upper limit on the cross section.

- *4-top search in multilepton final states:* As detailed in Chapter 1, 4-top searches target a rare process that is sensitive to many beyond-SM scenarios that involve third-generation quarks. One of the most powerful signatures to search for 4-top events involves the presence of multiple leptons with additional b-jets, which can also be produced by pair-produced $S_3^{1/3}$ Leptoquarks that decay into $t\tau$ and/or $t\mu$. The CMS Collaboration has recently performed a search for SM 4-top production in multilepton finals states (only electrons and muons are considered) using 35.9 $fb^{-1}$ of pp collision data at $\sqrt{s} = 13$ TeV [118]. Up to eight signal regions are defined depending on the number of light leptons (same sign dilepton or trileptons) and the number of b-tagged jets (2, 3, and $\geqslant 4$). Different requirements on the number of jets are made in different signal regions. In order to recast this search, a detailed study has been performed attempting to reproduce the experimental selection and the statistical analysis.

  We have implemented our model using `Feynrules` [14] and loaded it into `MadGraph5_aMC@NLO` [12], which has been used to simulate $S_3^{1/3}$ pair production at $\sqrt{s} = 13$ TeV, followed by the $S_3^{1/3}$ decay into both $t\tau$ and $t\mu$. Since we are focusing on pair production with the decay structure detailed in Eqs. 22-24, our results are in principle independent of the absolute value of the Leptoquark couplings to quarks and leptons. At most, their effects comes from the Leptoquark width, which is taken to be 1% of the mass, and thus smaller than the experimental resolution. The signal events have been generated using a leading-

order matrix element and the NN23LO1 PDF set [121], and have been processed through `Pythia 8` [16] for the modeling of parton showering and hadronization, as well as through a simulation of the CMS detector response as implemented in `Delphes 3` [18]. The simulated signal samples have been normalized using the NLO cross sections computed in Ref. [122] for squark pair production correcting for the fact that we do not have the same mass degeneration when considering Leptoquark pair production.

For a more accurate recasting of this search, we have modified the `Delphes 3` card and our data analysis to match as closely as possible the CMS detector and analysis described in Ref. [118]. With this modification we obtain simulated 4-top yields in each signal region that are in an average 15% disagreement with the central values reported in Ref. [118]. We scan the parameter space of Branching Ratios in our model, allowing for $S_3^{1/3}$ decays into $t\tau$, $t\mu$, and $b\nu$, and obtain the predicted yields in each signal region for this analysis. The slight discrepancies that had been found in the validation of 4-top yields are used to correct the yields per signal region. We derive upper limits on the signal cross section times Branching Ratio using the $CL_s$ method [123], which employs as test statistic the ratio of the likelihoods under the signal plus background hypothesis over the background-only hypothesis. The likelihood fits are performed using the `HistFitter` package [124], which relies on `RooFit` [125] and the minimization algorithms from `MINUIT` [126].

### 2.2.2 *Results*

In this Subsection we use the results from the LHC searches discussed in Subsection 2.2.1 to derive bounds in the plane of $BR(S_3^{1/3} \to t\tau)$ vs $BR(S_3^{1/3} \to t\mu)$ as a function of $S_3^{1/3}$ mass, under the assumption that $BR(S_3^{1/3} \to t\tau) + BR(S_3^{1/3} \to t\mu) + BR(S_3^{1/3} \to b\nu) = 1$. We also perform a simple extrapolation of these searches to higher energy $\sqrt{s} = 14$ TeV and luminosity $\mathcal{L} = 300$ fb$^{-1}$ which are representative limits of the full LHC Run 3 dataset. These bounds are compared to the preferred region within our model, given by Eqs. 32 and 33.

The three LHC searches that we study probe different final states and thus exhibit complementary sensitivity to the Branching Ratio parameter space, as shown in Fig. 3 for different values of $M_{S_3^{1/3}}$. The direct Leptoquark search [116] explores the $t\bar{t}\tau^+\tau^-$ final state and thus is sensitive to large $BR(S_3^{1/3} \to t\tau)$, while the sbottom search [115], focused on the $b\bar{b} + \text{MET}$ final state, probes large $BR(S_3^{1/3} \to b\nu)$. In the framework of partial compositeness these represent the two dominant decay modes for the $S_3^{1/3}$ Leptoquark. In contrast, the multilepton final states covered by the 4-top search [118] can originate from $t\bar{t}\mu^+\mu^-$, $t\bar{t}\tau^\pm\mu^\mp$, and $t\bar{t}\tau^+\tau^-$. Therefore, the 4-top search provides unique sensitivity to a large generation mixing $y_{32}$ which, although is not preferred by partial compositeness, is one of the parameters directly involved in the B-physics anomalies. It is worth recalling that this presentation of the results, where the bottom left corner of the Branching Ratio plane corresponds to large $BR(S_3^{1/3} \to b\nu)$, relies on the assumption of only three sizable decay modes ($S_3^{1/3} \to t\tau, t\mu, b\nu$), and would be modified if additional sizable unknown decay modes would exist. Furthermore, it should be stressed that the parameter space being probed is a function of the Branching Ratios and not the absolute values of the weak couplings. Therefore, since

$M_{S_3^{1/3}} >> m_t$ yields Branching Ratios independent of $M_{S_3^{1/3}}$, practically the whole parameter space as function of Branching Ratios in Fig. 3 is in principle allowed by the B-physics anomalies, namely Eq. 25. The extreme cases of Branching Ratios equal to zero, one, or very close to them, may yield $y_{ij}$ that either cannot satisfy Eq. 25 or have non-perturbative values.

Based on these results, the partial compositeness preferred region is excluded for masses $M_{S_3^{1/3}} < 0.9$ TeV. For higher masses, the region preferred by our model is not directly excluded. This conclusion assumes that searches for the other components of the multiplet -namely $S_3^{4/3}$ and $S_3^{2/3}$- do not affect the $S_3^{1/3}$ component. This depends on the mass splitting $\Delta m_S \sim \mathcal{O}(100\ \text{GeV})$ and the sensitivity of these other searches. In particular, the $S_3^{-2/3}$ Leptoquark decays with 100% into $t\nu$ and thus is constrained by stop searches in the $t\bar{t} + \text{MET}$ final state, which yield a mass limit $M_{S_3^{2/3}} > 1.07$ TeV [115]. In contrast, the $S_3^{4/3}$ Leptoquark can be sought directly [113] or probed by the $B - L$ R-parity-violating stop search in Ref. [114]. The final state in this search is the same as in $S_3^{4/3}$ pair-production except for a $b$-$\bar{b}$ distinction which becomes irrelevant at the detector level. The actual lower mass limit on $M_{S_3^{4/3}}$ depends on its Branching Ratios. For a large Branching Ratio into $b\tau$ the limit is ~0.9 TeV, whereas for large Branching Ratio into $b\mu$ the limit reaches ~1.4 TeV.

It is interesting to project the bounds on the $S_3^{1/3}$ parameter space for upcoming LHC runs. We use the following procedure to estimate the projected sensitivities for the full Run 3 dataset ($\sqrt{s} = 14$ TeV, 300 fb$^{-1}$) of the relevant searches discussed above. We obtain the signal cross-section at $\sqrt{s} = 14$ TeV from the NLO calculation for squark pair production in Ref. [122]. For the direct Leptoquark search in the $t\bar{t}\tau^+\tau^-$ channel we assume that background yields scale with a common factor of 1.2 going from $\sqrt{s} = 13$ TeV to 14 TeV, and that the analysis is statistically limited and thus background uncertainties scale accordingly with the integrated luminosity. For the sbottom search we use a similar procedure and we obtain results consistent with those in Ref. [127]. To project the limits from the 4-top search we rescale the estimated backgrounds to their NLO cross section at $\sqrt{s} = 14$ TeV obtained from `MadGraph5_aMC@NLO` [12], and we perform the statistical analysis based on the expected signal and background yields.

The projected exclusion regions are shown in Fig. 4, which can be compared to the current ones in Fig. 3. Using the full Run 3 dataset, the full Branching Ratio plane would be excluded for masses $\leqslant 0.9$ TeV and a significant portion of the parameter space can be probed for a mass of ~ 1 TeV. To better visualize the LHC reach we display in Fig. 5 the exclusion regions under the assumption that $\text{BR}(S_3^{1/3} \to t\mu)$ is negligible as expected in partial compositeness. In this case we consider only the searches in the $t\bar{t}\tau^+\tau^-$ and $b\bar{b} + \text{MET}$ final states. In this scenario, the Branching Ratios are fixed to $\text{BR}(S_3^{1/3} \to t\tau) = \text{BR}(S_3^{1/3} \to b\nu) = 0.5$ and masses up to 1.2 TeV can be probed, driven by the $b\bar{b} + \text{MET}$ search.

We proposed in Ref. [102] that a dedicated search targeting the $t\tau b\nu$ final state, which has the largest Branching Ratio, is relevant to probe masses somewhat higher than 1.2 TeV. Preliminary results on this final state have been reported by the CMS collaboration in Ref. [128], where they exclude pair produced scalar Leptoquarks of masses $\leqslant 0.95$ TeV for $\text{BR}(S_3^{1/3} \to t\tau) = \text{BR}(S_3^{1/3} \to b\nu) = 0.5$. In any case, exploring beyond the 1–1.5 TeV mass scale for the $S_3^{1/3}$ Leptoquark will require pursuing different strategies, including considering single- and non-resonant production.

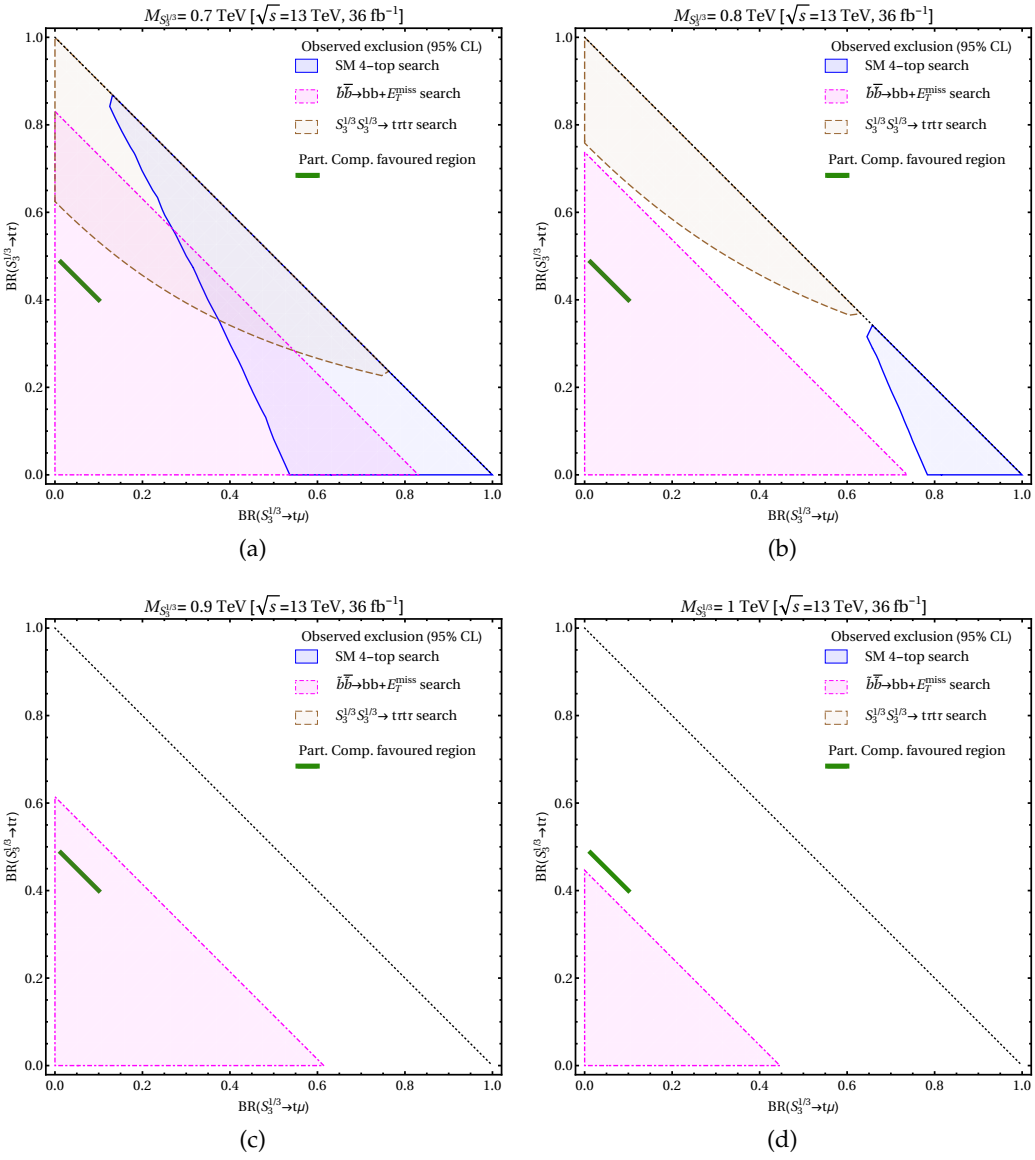



Figure 3: Bounds on the $S_3^{1/3}$ Leptoquark Branching Ratios into tτ, tμ, and bν, derived from recent LHC searches based on 36 fb$^{-1}$ of pp collisions at $\sqrt{s} = 13$ TeV, for different values of the Leptoquark mass $M_{S_3^{1/3}}$: (a) 0.7 TeV, (b) 0.8 TeV, (c) 0.9 TeV, and (d) 1 TeV. Also shown is the region preferred by our partial compositeness model. The area above the diagonal line corresponds to the unphysical region where the sum of Branching Ratios exceeds unity, or is smaller than zero.

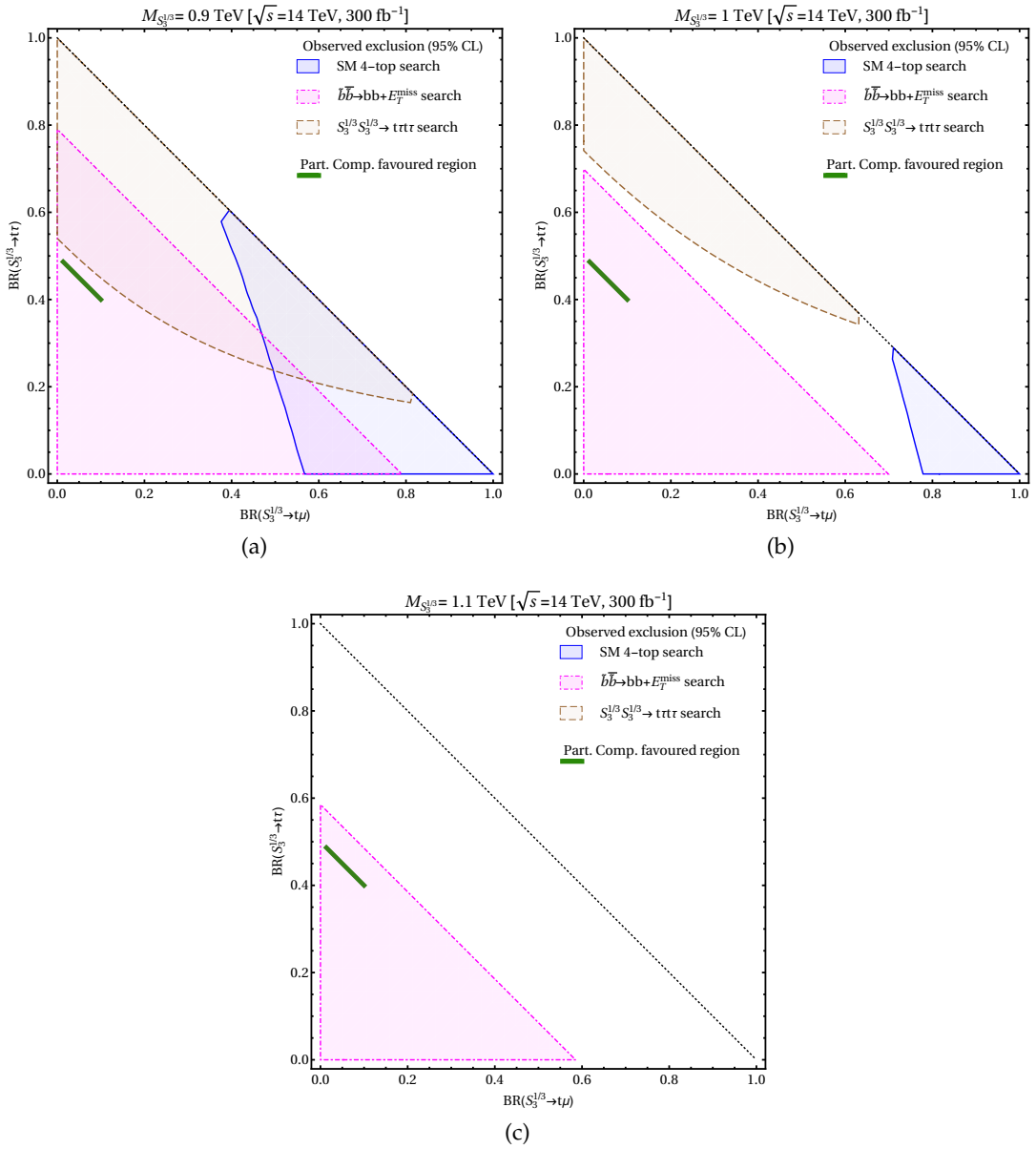



Figure 4: Projected bounds on the $S_3^{1/3}$ Leptoquark Branching Ratios into tτ, tμ, and bν based on recent LHC searches extrapolated to 300 fb$^{-1}$ of pp collisions at $\sqrt{s} = 14$ TeV, for different values of the Leptoquark mass $M_{S_3^{1/3}}$: (a) 0.9 TeV, (b) 1 TeV, and (c) 1.1 TeV. Also shown is the region preferred by our partial compositeness model. The area above the diagonal line corresponds to the unphysical region where the sum of Branching Ratios exceeds unity, or is smaller than zero.

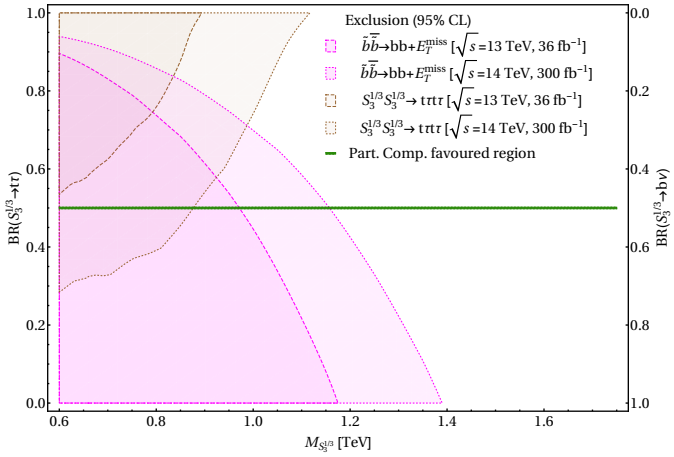



Figure 5: Bounds on the $S_3^{1/3}$ Leptoquark Branching Ratios into $t\tau$ and $b\nu$ (assuming a negligible $t\mu$ Branching Ratio) as a function of Leptoquark mass $M_{S_3^{1/3}}$. The current bounds (dashed line) and projected bounds (dotted line) for the LHC Run 3 derived from the $t\bar{t}\tau^+\tau^-$ and $b\bar{b}+\text{MET}$ searches are compared. Also shown is the region preferred by our partial compositeness model.

## 2.3 DISCUSSION

In this Chapter we use the results from LHC searches to constrain the parameter space of a model with a scalar Leptoquark $S_3$ that transforms as $(\bar{3},3)_{1/3}$. This model is motivated by the B-physics anomalies and encloses the $S_3$ state and the SM Higgs boson as pNGBs of a strongly interacting sector at higher energies. The pNGB scenario allows to justify why this scalar particle would be the lightest new state in the spectrum and thus potentially within reach at the LHC, whereas other fermionic and vector resonances would be heavier and potentially beyond kinematic reach. We also exploit the relationships between couplings of the Leptoquark to quarks and leptons of different generations that arise within the framework of partial compositeness.

We study the production of a pair of Leptoquarks, $pp \to S_3\bar{S}_3$, since its cross section does not depend on its couplings to leptons and quarks, and it should represent the dominant production mechanism for a Leptoquark in the mass regime currently probed at the LHC (up to $\sim$ 1 TeV) for small enough couplings. The different decay modes by each of the three $S_3$ components result in different search strategies.

The $S_3^{2/3}$ Leptoquark is the most straightforward one to probe since its three decay channels all lead to the same final state signature, $S_3^{-2/3}S_3^{+2/3} \to t\bar{t}+\text{MET}$. Consequently, current stop searches can be effectively used to exclude $S_3^{-2/3}$ masses below $\sim$1.1 GeV [115].

In contrast, pair production of the $S_3^{4/3}$ Leptoquark results in final states with pairs of opposite-charge leptons plus b-quark jets, $S_3^{-4/3}S_3^{+4/3} \to b\ell\bar{b}\ell'$ ($\ell,\ell' = e,\mu,\tau$), although there is a prejudice towards same-flavour dilepton final states with the lepton being $\mu$ or $\tau$. Lower mass limits for this Leptoquark come from dedicated searches [113] and from constraints in $B-L$ R-parity violating stop searches [114] and, depending on the assumed Branching Ratio, can range from $\sim$0.9 GeV (mainly $b\tau$) to $\sim$1.4 GeV (mainly $b\mu$).

Finally, the $S_3^{1/3}$ Leptoquark has six decay channels, giving multiple final states that can be probed at the LHC: $S_3^{+1/3}S_3^{-1/3} \to t\ell\bar{t}\ell'$, $b\bar{b} + \text{MET}$, and $t\ell b + \text{MET}$. Therefore, the LHC is in principle more sensitive to probe this Leptoquark.

Since we can expect within our model a mass splitting between the components of the $S_3$ multiplet of up to about $\mathcal{O}(100\,\text{GeV})$, applying lower mass limits derived for one component to the other components is not direct. In this Chapter we have mainly focused on constraining the parameter space of the $S_3^{1/3}$ component. Further work within the proposed model could relate these constraints to limits in the other components.

Within our model, the main decay channels of the $S_3^{1/3}$ Leptoquark are $S_3^{1/3} \to t\tau, t\mu, b\nu$. We study three complementary searches at the LHC using 36 $\text{fb}^{-1}$ at $\sqrt{s} = 13$ TeV that can constrain the corresponding Branching Ratio parameter space. These include dedicated searches for $t\bar{t}\tau^+\tau^-$ [116] and $b\bar{b} + \text{MET}$ [115, 120], which are primarily sensitive to large $\text{BR}(S_3^{1/3} \to t\tau)$ and $\text{BR}(S_3^{1/3} \to b\nu)$, respectively. In addition, we recast a search for SM 4-top production in multilepton final states [118], which is sensitive to large $\text{BR}(S_3^{1/3} \to t\mu)$.

We find that most of the Branching Ratio parameter space is excluded up to masses of about ~800 GeV, and that the preferred region for our model, $\text{BR}(S_3^{1/3} \to t\tau) \simeq \text{BR}(S_3^{1/3} \to b\nu) \simeq 0.5$, is best probed by the $b\bar{b} + \text{MET}$ search and remains allowed for masses above 1 TeV. We also project the sensitivity of these searches to the full LHC Run 3 dataset (300 $\text{fb}^{-1}$ at $\sqrt{s} = 14$ TeV). We find that for a dominant Branching Ratio to $t\tau$ ($b\nu$) it will be possible to exclude masses up to 1.1 TeV (1.4 TeV), whereas for the more realistic scenario of equally split Branching Ratios, as predicted by our model, the expected limit is about 1.2 TeV. A proposed dedicated search targeting the $t\tau b\nu$ final state was implemented in Ref. [128].

Finally, we appraise that for mass scales well above 1–1.5 TeV, and for regions of parameter space consistent with the B-physics anomalies, single- and non-resonant production are expected to dominate, requiring optimized search strategies.

# 3

# LEPTOQUARK SINGLE- AND NON-RESONANT PRODUCTION AT THE LHC

In this Chapter we study a moderate excess reported by CMS in Ref. [129]. This excess is discarded based on arguments which are valid only for a broad but still specific modelling of Leptoquark production. We view this as an opportunity to reconsider the assumptions made when searching for Leptoquarks and check whether a departure from these assumptions could in principle yield a signal like the one discarded. To this goal, we implement a purely phenomenological but reasonable inter-generational Leptoquark model where single- and non-resonant production as defined in Subsection 1.2.1 are also taken into account. This model provides the adequate features for a non-standard signal and could be taken as a starting point towards modifying search strategies to account for a larger set of possible models. This Chapter is based on work done in collaboration with Ezequiel Alvarez in Ref. [130].

This Chapter is organized as follows. In Section 3.1 we detail the aforementioned search and the characteristics of the excess that causes the collaboration to dismiss its possible New Physics (NP) origin. In Section 3.2 we establish a purely phenomenological Leptoquark model which avoids LHC constraints. This arbitrary model is tailored to yield the necessary signal features to explain the moderate excess while avoiding some of the arguments which are used to dismiss it. This is shown in Section 3.4. In Section 3.5 we consider low-energy constraints and a how to avoid them. We discuss deviations from our hypotheses, the $\mu\mu jj$ defect and interference effects in Section 3.6. Finally, we summarize and propose different experimental handles to tackle future searches in Section 3.7.

## 3.1 THE CMS RESULT

The CMS collaboration reported in Ref. [129] a search for second generation Leptoquark pair production in the $\mu\mu jj$ and $\mu\nu jj$ channels. Both channels require at least two jets with $p_T > 50$ GeV, with these two leading jets taken as coming from the decay of two Leptoquark candidates, and at least two (exactly one) muons with $p_T > 53$ GeV for the $\mu\mu jj$ ($\mu\nu jj$) channel. The two Leptoquark candidates are determined by pairing the two leading jets with the two leading muons or the muon and $p_T^{\text{miss}}$. The correct pairing is assumed to be the one that minimizes the (transverse) mass difference between the two Leptoquark candidates for the $\mu\mu jj$ ($\mu\nu jj$) channels. After event selection and reconstruction, discrimination against SM backgrounds for each channel is optimized using three kinematic variables: $S_T^{\mu\mu jj}$, $m_{\mu\mu}$ and $m_{\mu j}^{min}$ for $\mu\mu jj$ and $S_T^{\mu\nu jj}$, $m_{\mu\nu}^T$ and $m_{\mu j}$ for $\mu\nu jj$, where $m_{\mu j}^{min}$ and $m_{\mu j}$ are calculated only on the paired muons and jets. $m_{ab}$ ($m_{ab}^T$) is the (transverse) invariant mass of the particles $a$ and $b$ while $S_T^{abcd}$ is the transverse momentum scalar sum of particles $a, b, c$ and $d$, as shown in Eq. 34:

$$\begin{aligned} m_{ab} &= \sqrt{E_a E_b - \vec{p}_a \cdot \vec{p}_b} \\ m_{ab}^T &= \sqrt{2 p_T^a p_T^b (1 - \cos\Delta\phi_{ab})} \\ S_T^{abcd} &= \sum_{i=a,b,c,d} |\vec{p}_T^{\,i}| \end{aligned} \tag{34}$$

$m_{\mu j}^{min}$ is the minimum of the two masses obtained after pairing muons and jets in $\mu\mu jj$. Different cuts in these kinematic variables define different signal regions and each optimized signal region corresponds to a Leptoquark mass hypothesis. The Leptoquark mass hypothesis for which the cuts are optimized is varied between 200 GeV and 2000 GeV. Since each kinematic variable is a monotonically growing function of the mass of the candidate Leptoquark (see Fig. 4 of Ref. [129]), events contained in a given mass signal region form a subset of the events contained in a lower mass signal regions. To define each signal region that is optimized for a Leptoquark of given mass $M_{S^{1/3}}$, Ref. [129] uses the variable $M_{LQ}$. It should be stressed that $M_{LQ}$ is not a really a mass, but a signal region optimized for a Leptoquark of mass $M_{S^{1/3}} = M_{LQ}$. Further details on these signal regions can be found in the CMS paper Ref. [129].

The CMS collaboration reports event yields in each signal region and there is a qualitative different result for the observed versus expected number of events in each final state. In the $\mu\mu jj$ channel, there is a deficit for the signal regions $M_{LQ} \sim 600\ \text{GeV} - 800\ \text{GeV}$, with a significance in a specific bin of about $2 \sim 3$ standard deviations and an anomalous event in the high $S_T^{\mu\mu jj}$ region where no SM events are expected. The deficit is not discussed while the anomalous event is discarded due to the two Leptoquark masses candidates not compatible with pair production. In the $\mu\nu jj$ channel, there is a reported excess in the $M_{LQ} \sim 900\ \text{GeV} - 1100\ \text{GeV}$ signal regions. This excess is greater at $M_{LQ} = 950$ GeV, where the expected background events is $16.9 \pm 1.0 \pm 1.7$ and the measured data is 30 events [129]. Including a poissonian uncertainty for the data and adding all uncertainties in quadrature yields a significance of 2.25 standard deviations. Because of the correlation between each signal region $M_{LQ}$ it is not possible at this level to perform a multi-$M_{LQ}$ analysis for the deficit or the excess. In the CMS paper [129] it is stated that the given excess does not show a peak in the $m_{\mu j}$ distribution, which is characteristic of Leptoquark pair production since the final particles would come from on-shell NP particles.

In the following sections we aim to reproduce the excess in the $\mu\nu jj$ final state without significantly altering $\mu\mu jj$. This excess also needs to correspond to a mass distribution without the resonant structure expected from pair production. We discuss the deficit in $\mu\mu jj$ and the anomalous event in Section 3.6.

### 3.2 A SIMPLE LEPTOQUARK MODEL

For simplicity, we assume that there is only one species of Leptoquark involved in the production of the $\mu\mu jj$ and $\mu\nu jj$ final states. If a given Leptoquark decays to $\mu j$ and also $\nu j$ then its electric charge can be either 1/3 or 2/3. As we show in Section 3.4, the former is more suited to explain the excess in the $\mu\nu jj$ final state without providing an excessive yield in the $\mu\mu jj$ final state. Therefore we assume a Leptoquark with $Q = 1/3$.

The main objective in this Chapter is to address the results in CMS paper Ref. [129]. Just for the sake of concreteness, we work with a Leptoquark that is also involved

in a possible solution for the B-anomalies, as for instance $S_3$ or $S_1$ [39, 45–51]. We choose to work with the $S_1$ Leptoquark with quantum numbers $S_1 \sim (\bar{3}, 1)_{1/3}$ defined in Section 1.2.1. Being a $SU(2)$ singlet, $S_1$ does not present constraints coming from other multiplet partners of different charge. As detailed in Subsection 1.2.1, its quark-lepton interaction Lagrangian can be written as [35]

$$\mathcal{L} \supset y^{LL}_{ij} \bar{Q}^{C\,i,a}_{L} S_1 \, \epsilon^{ab} \, \ell^{j,b}_{L} + y^{RR}_{ij} \bar{u}^{C\,i}_{R} S_1 \, e^{j}_{R} + y^{\overline{RR}}_{ij} \bar{d}^{C\,i}_{R} S_1 \, \nu^{j}_{R}. \tag{35}$$

Where $\epsilon$ is the anti-symmetric tensor and $y^{LL,RR,\overline{RR}}_{ij}$ are the NP couplings constants for left-handed, up-type right-handed and down-type right-handed quark-lepton currents respectively. Neglecting diquark interactions, this and the QCD interactions encoded in the kinematic Lagrangian are the only relevant interactions.

Throughout this Chapter we take the following ansatz for the couplings at the TeV scale

$$y^{LL} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & y_{22} & y_{23} \\ 0 & y_{32} & y_{33} \end{pmatrix} \tag{36}$$

and $y^{RR}_{ij} = y^{\overline{RR}}_{ij} = 0$.

Although we present this model from a phenomenological point of view in order to perform a proof of concept on the CMS paper Ref. [129], one may observe that the pattern of a Leptoquark coupled mainly to second and third generations at the TeV scale can be expected in models where Leptoquarks are pseudo-Nambu Goldstone Bosons of a strongly interacting heavy sector as in Chapter 2. For simplicity, we take all $y_{ij}$ to be real and non-negative. However, they could have different signs without affecting this analysis.

The ansatz in Eq. 36 avoids first generation constraints by forbidding Leptoquark interactions with electrons and certain flavour constraints by forbidding right-handed currents [35, 131]. As shown in Section 3.4, to account for the reported results in [129] Leptoquark single- and non-resonant production have to be competitive with pair production. For TeV scale masses, this can be achieved by assuming some components in $y_{ij}$ are of $\mathcal{O}(1)$ [51, 119]. For simplicity we assume in the following that

$$y_{23},\, y_{32} \ll 1. \tag{37}$$

Since along the text only $y_{22,33}$ are non negligible, when referring to a Leptoquark coupling to second or third generation we take it to mean that the Leptoquark is diagonally coupled to both lepton and quark of the given generation. The case where $y_{32}$ and $y_{23}$ are non negligible is discussed in Section 3.6.

With these assumptions and the Lagrangian in Eq. 35, this particle has the following open decay channels:

$$S^{-1/3} \to c\,\mu^-,\; t\,\tau^-,\; s\,\nu_i,\; b\,\nu_i. \tag{38}$$

and the Leptoquark width is determined by

$$\begin{aligned} \Gamma_{S^{1/3}} &= \sum_{q_i = u_i, d_i} \sum_{L_j = \ell_j, \nu_j} \frac{|y_{q_i L_j}|^2}{16\pi M^3_{S^{1/3}}} (M^2_{S^{1/3}} - m^2_{q_i} - m^2_{L_j}) \\ &\quad \sqrt{M^4_{S^{1/3}} - 2M^2_{S^{1/3}}(m^2_{q_i} + m^2_{L_j}) + (m^2_{q_i} - m^2_{L_j})^2} \\ |y_{u_i \ell_j}| &= (V^T y^{LL*})_{ij} \\ |y_{d_i \nu_j}| &= (y^{LL} U)_{ij} \end{aligned} \tag{39}$$

where V and U are the CKM and PMNS matrices. Assuming $V_{CKM} \approx 1$, using the unitarity of the neutrino mixing matrix $U_{PMNS}$, and assuming a top quark mass negligible compared to $M_{S^{1/3}}$, one can derive the Branching Ratios

$$\begin{aligned} BR(S^{-1/3} \to c\mu^-) &= \textstyle\sum_i BR(S^{-1/3} \to s\nu_i) = \frac{y_{22}^2}{2(y_{22}^2+y_{33}^2)} \\ BR(S^{-1/3} \to t\tau^-) &= \textstyle\sum_i BR(S^{-1/3} \to b\nu_i) = \frac{y_{33}^2}{2(y_{22}^2+y_{33}^2)} \end{aligned} \tag{40}$$

If one is agnostic to jet flavour and assumes $y_{22} \approx y_{33}$, the Branching Ratios obey that

$$\sum_i BR(j\nu_i) \approx 2\, BR(j\mu) \tag{41}$$

The difference between neutrinos and muons resides in that neutrinos are produced with both second and third generation jets due to $y_{33} \neq 0$ while muons are produced only with second generation c-jets. As we state at the end of this Section, this difference between neutrinos and muons can help understand the differences between the $\mu\mu jj$ and $\mu\nu jj$ channels in the CMS paper Ref. [129].

To study this interplay between the two generations relevant to $j\nu$ decays, it is useful to define the parameter

$$r = \frac{y_{33}}{y_{22}}. \tag{42}$$

The Branching Ratios can then be written as

$$\begin{aligned} BR(S^{-1/3} \to c\mu^-) &= \textstyle\sum_i BR(S^{-1/3} \to s\nu_i) = \frac{1}{2(1+r^2)} \\ BR(S^{-1/3} \to t\tau^-) &= \textstyle\sum_i BR(S^{-1/3} \to b\nu_i) = \frac{r^2}{2(1+r^2)} \end{aligned} \tag{43}$$

The Branching Ratio to second and third generation are plotted as a function of r in Fig. 6. For $r < 1$, the second generation is dominant while for $r > 1$ the third generation dominates.

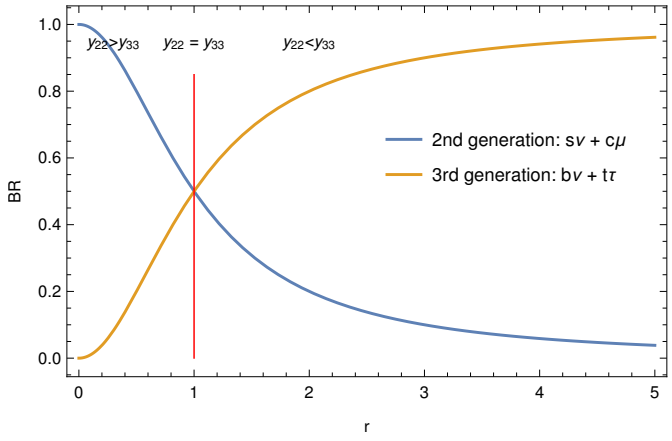



Figure 6: Branching Ratio for the second and third generations as a function of the parameter r (see Eq. 42). The Branching Ratios to each specific channel is the half of the plotted Branching Ratio to the corresponding generation.

As discussed in Sections 3.3 and 3.4, when the third generation Branching Ratio dominates, the model can avoid tensions in purely second generation final states such as $\mu\mu jj$ without sacrificing a larger event yield in channels sensitive to third generation

couplings such as μνjj. This corresponds to the parameter region where $y_{22} \leqslant y_{33}$ which implies $r \geqslant 1$.

However, we also need to determine the absolute value of the couplings and the mass. Here it is important to remember that we need not only a difference between channels but also a non-typical kinematic distribution. For any Leptoquark model, the smaller the Leptoquark mass $M_{S^{1/3}}$ and the couplings $y_{ij}$, the more favored is pair production whereas the larger they are, the more favored is single- and non-resonant production [51, 119]. Since the CMS paper Ref. [129] observes that kinematic distributions in the deviation lack the characteristic mass peak expected from pair produced on-shell Leptoquarks, it is convenient to explore a region in parameter space where μνjj and μμjj single- and non-resonant production could hide on-shell characteristic kinematic distributions. We find that the region determined by

$$\begin{aligned} M_{S^{1/3}} &\sim \mathcal{O}(1\ \text{TeV}) \\ y_{22},\, y_{33} &\sim \mathcal{O}(1) \end{aligned} \tag{44}$$

successfully achieves the above requirements, while keeping the physics perturbative and within the scope of current LHC analyses. We plot in Fig. 7 the size of the different contributions to the total cross-section (with basic cuts as outlined in Ref. [129]) of both μμjj and μνjj final states as a function of the coupling $y_{22}$ for fixed r and Leptoquark mass which are justified in Section 3.3. We find that for $y_{22} \sim 0.4$ pair production becomes sub-leading.

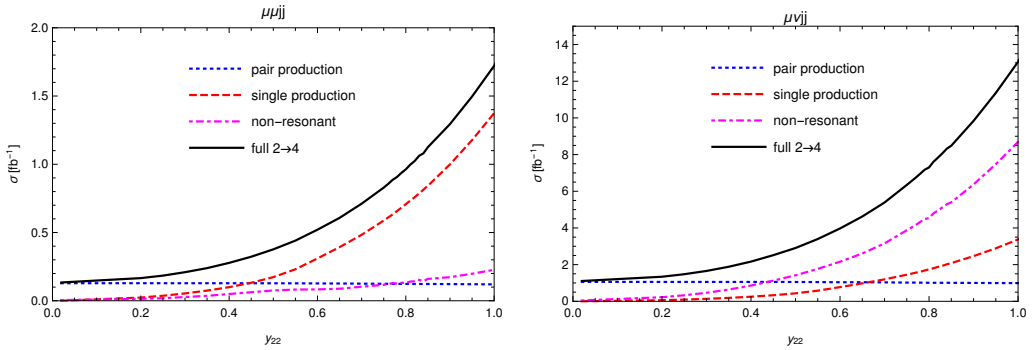



Figure 7: Pair, single- and non-resonant production contributions to the final states μμjj and μνjj as a function of $y_{22}$. Other parameters are fixed at $r = 1.7$ and $M_{S^{1/3}} = 950$ GeV. In both cases for $y_{22} \sim 0.4$ resonant pair production becomes sub-leading. Due to the relative size of the couplings for each generation, $y_{33}/y_{22} = r = 1.7$, the non-resonant contribution is the dominant for μνjj, whereas single-resonant production contribution dominates for μμjj.

Before advancing to the LHC constraints and the choice of an specific benchmark point to test against the CMS result, we delve further into the production mechanisms for the final states μμjj and μνjj in which $S^{1/3}$ has a relevant contribution. We do this to qualitatively understand how the presented model can yield larger contributions to μνjj than to μμjj to better agree with the results in the CMS paper Ref. [129]. There are two relevant differences between those two final states: *i)* Leptoquark Branching Ratios (relevant to pair production and decay), and *ii)* quark abundances in the proton pdf (relevant to single- and non-resonant production), with only the latter difference depending on the chosen $M_{S^{1/3}}$ and absolute value of the couplings.

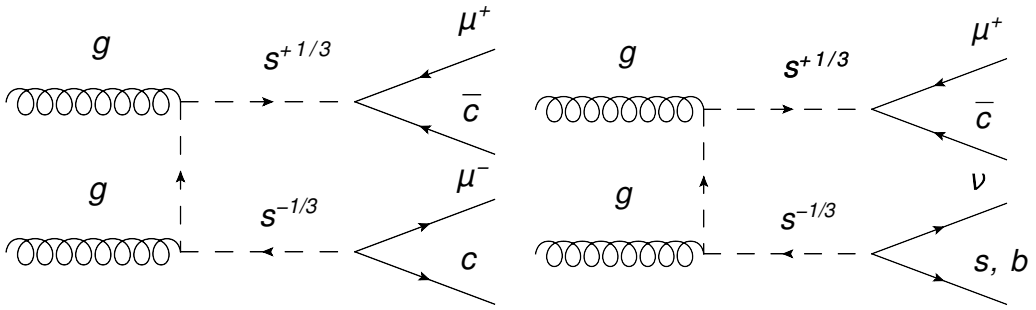



Figure 8: Representative QCD pair production mechanism diagrams for the final states a) $\mu\mu jj$ and b) $\mu\nu jj$. As $r$ increases the $\mu\nu jj$ final state is considerably preferred over $\mu\mu jj$.

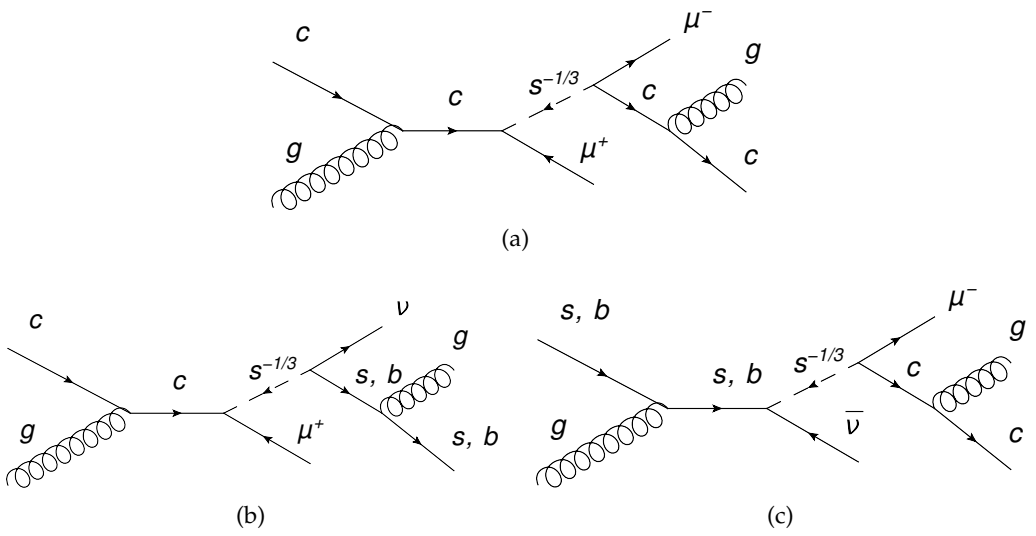



Figure 9: Representative single-resonant production mechanism diagrams for the final states a) $\mu\mu jj$ and b) and c) $\mu\nu jj$. Notice how c g initial state contributes to $\mu\mu jj$ and $\mu\nu jj$, whereas b/s g initial state only contributes to $\mu\nu jj$.

The relevant processes for pair production of $\mu\mu jj$ and $\mu\nu jj$ are shown in Fig. 8, where we see that the production process depends on the QCD gauge coupling between gluons and Leptoquarks and not on the $y_{ij}$ couplings. The cross-section for processes involving pair production can be approximated as

$$\begin{aligned}\sigma(gg \to (S^{1/3} \to l_1 q_1)(S^{-1/3} \to l_2 q_2)) \quad &\simeq \quad \sigma(gg \to S^{1/3} S^{-1/3}) BR(S^{1/3} \to l_1 q_1) \\ &\quad\; BR(S^{-1/3} \to l_2 q_2).\end{aligned}$$

and thus the only relevant information from the couplings is their ratio $r$. Because of this, the Branching Ratios alone yield the necessary difference between the $\mu\mu jj$ and $\mu\nu jj$ channels.

When $y_{22} \approx y_{33}$ one sees from Eq. 40 that $\mu\nu jj$ pair production is approximately four times larger than $\mu\mu jj$ pair production. The reason is that due to neutrino mixing $y_{33} \neq 0$ opens a channel to third generation decays which enhance the cross-section. This effect alone could explain the observed difference between channels, but, as it is reported in [129], the absence of a peak in the distributions indicates that single-

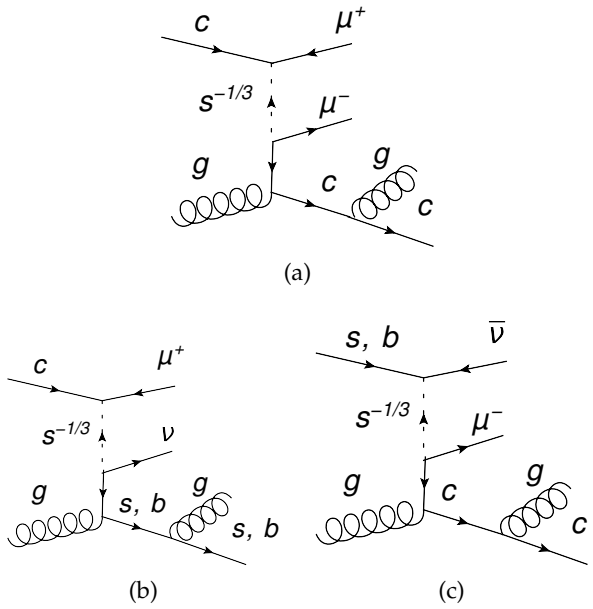



Figure 10: Representative non-resonant production mechanism diagrams for the final states a) $\mu\mu jj$ and b) and c) $\mu\nu jj$. As in single-resonant production, notice that c g initial state contributes to $\mu\mu jj$ and $\mu\nu jj$, whereas $b/s$ g initial state only contributes to $\mu\nu jj$.

and non-resonant production should also be taken into account. In pair production all final states come from on-shell particles, whereas this is not valid for single- and non-resonant production. In these channels an argument relying exclusively on the Branching Ratios no longer applies and one should consider different arguments.

In contrast to pair production which is produced from a gg initial state, single- and non-resonant productions require quarks in the initial state. This generates a new imbalance that favors $\mu\nu jj$ over $\mu\mu jj$. In Figs. 9 and 10 we show some representative Feynman diagrams for the single- and non-resonant production mechanisms. We can see in those Figures how there is a difference between the partonic processes relevant in each channel. We refer in the following paragraphs to partonic processes with disregard to whether the quarks and leptons involved are particles or antiparticles, a distinction which does not matter to our argument since we are not considering first generation quarks. There is an imbalance when comparing

$$s\,g \to \mu\nu jj \qquad \text{vs.} \qquad c\,g \to \mu\mu jj \tag{45}$$

because the pdf abundance of the s-quark is larger than the c-quark. Moreover, when analyzing the analogous diagrams but at the third generation level in a 5-flavour scheme the process

$$b\,g \to \mu\nu jj \tag{46}$$

does not have its counterpart which would require top-quark in the proton PDF. This imbalance between up-type and down-type quarks enhances $\mu\nu jj$ over $\mu\mu jj$.

There are also other sources of imbalance between the $\mu\mu jj$ and $\mu\nu jj$ channels. In single- and non-resonant production, the initial state c g contributes to the final state $\mu\nu jj$ channel whereas the $s/b$ g initial state does not have it corresponding diagram contributing to the final state $\mu\mu jj$. This is exemplified by the single-resonant repre-

sentative Feynman diagrams in Fig. 9. This effect is not due to quark abundances in the proton but to the definition of the channels.

The above paragraphs indicate how one can expect a larger deviation in $\mu\nu jj$ than in $\mu\mu jj$ for many production processes. This difference increases even more if $r = y_{33}/y_{22}$ increases. For pair production because the difference between third and second generation Branching Ratios grows. And for single- and non-resonant production because b-associated diagrams have a larger coupling constant.

### 3.3 LHC CONSTRAINTS

Although, the region in parameter space defined in Eq. 44 yields single-, non-resonant and pair production within the same order of magnitude, the main direct bounds set by the LHC data come from pair production searches. In the following paragraphs we obtain the constraints coming from this kind of searches. However, one should be cautious in the limit setting procedure as the competitiveness of single- and non-resonant production would affect signal and background control regions needed to set the limits. Since we are considering a region in parameter space where single and non-resonant dominates over resonant pair production, we expect that NP affects more the background estimation than the signal counting. Therefore, assuming only on-shell pair production NP in a scenario where single- and non-resonant mechanisms may not be negligible would yield stronger limits than what they actually are, since backgrounds estimated through data-driven techniques would be under-estimated. Work in the direction of considering non-resonant effects in existing searches can be found in Ref. [132, 133].

The potentially problematic channels can be seen from Eq. 38 with the relevant Branching Ratios for $S^{-1/3}$ detailed in Eq. 43. They are the $t\tau^-$, $b + \text{MET}$, $s + \text{MET}$ and $c\mu^-$ channels. We investigate these four channels bearing in mind that the results on the latter are the ones reconsidered in this Chapter.

As the searches focus on pair production, they assume a Branching Ratio scheme to set an upper bound on the cross-section. $t\tau^-$ and $b + \text{MET}$ are recasted from third generation Leptoquark searches in the $t\bar{t}\tau^+\tau^-$ final state [116] and the supersymmetric search of sbottom pair production through the $b\bar{b} + \text{MET}$ final state [115, 120] for the case of zero neutralino mass, respectively. Limits on $s + \text{MET}$ final state are obtained from recasting the squark pair production search designed for first and second generation in final state $jj + \text{MET}$ [134]. To recast the limits found in these searches, we match the measured cross-section and the assumed Branching Ratio to the point in parameter space of our model. $c\mu^-$ is recasted from second generation Leptoquark searches [129, 135] which assume a $\text{BR}(S^{-1/3} \to c\mu^-)$ of either 1 or 1/2. Due to the chosen chiral couplings, $S^{1/3}$ has a maximum Branching Ratio of 1/2 to any decay channel. This weakens the limits set on $M_{S^{1/3}}$ for third generation searches as the maximum cross-section allowed can increase by a factor of at most 4.

If we parametrize the Branching Ratios with $r$, as in Eqs. 42 and 43, we can recast the limits set by each search to a limit on $M_{S^{1/3}}$ for each $r$ as can be seen in Fig. 11. As the third generation Branching Ratios increase with $r$, third generation searches set upper bounds on the allowed $r$. On the other hand, second generation Branching Ratios decrease with $r$ and second generation searches set a lower bound on $r$ for each $M_{S^{1/3}}$. Observe that since any Branching Ratio is bounded by 1/2, then all Leptoquark masses whose cross-section is less than four times the limit set by the searches are allowed for any value of $r$.

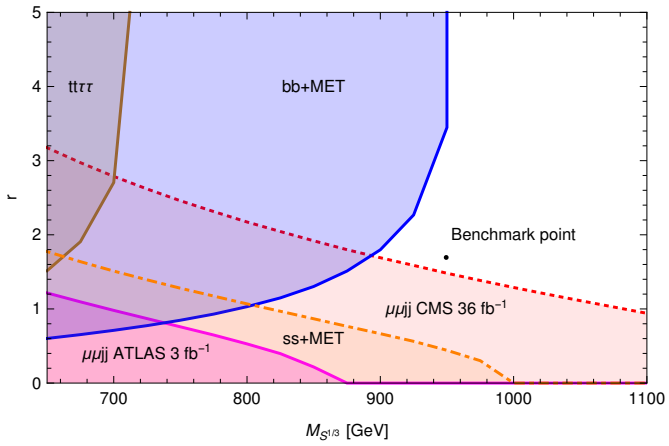



Figure 11: Excluded regions in the $M_{S^{1/3}}$ vs. $r$ plane. The dashed line represents the limits from CMS paper Ref. [129] assuming no NP effects. The $\mu\mu jj$ ATLAS limit is recasted from Ref. [135], the $tt\tau\tau$ limits is recasted from Ref. [116], the $bb + \text{MET}$ limit is recasted from Ref. [115], and the $ss + \text{MET}$ limits is recasted from Ref. [134].

Fig. 11 summarizes the above described analysis recasted as limits on the allowed values of $r$ for each possible $M_{S^{1/3}}$. We also indicate a benchmark point defined in the following paragraphs. As expected from the previous discussion, for $M_{S^{1/3}}$ large enough $r$ is no longer bounded.

### 3.3.1 *Benchmark point*

Following the excluded regions in Fig. 11 and motivated by the roughly two standard deviation excess in the CMS paper Ref. [129], we consider the benchmark point defined by

$$\begin{aligned} y_{33} &= 1.2 \\ y_{22} &= 0.7 \\ M_{S^{1/3}} &= 950 \text{ GeV}, \end{aligned} \tag{47}$$

and is indicated in Fig. 11.

The benchmark point yields a width of $\Gamma \sim 7.5\%$ of the mass, with Branching Ratios within the constraints on pair production from Refs. [115, 120], $\text{BR}(S^{-1/3} \to t\tau^-) = 0.37$ and $\text{BR}(S^{-1/3} \to c\mu^-) = 0.13$. It should be noted that a large $y_{33}$ allows the third generation to act as an escape valve against constraints from second generation Leptoquark searches and to populate the $\mu\nu jj$ final state.

It is worth noting that this benchmark point, having large couplings to enhance non-resonant production, is likely to yield potential problems in flavour physics [35, 46–51, 136] which should be addressed in detail. With this purpose, we perform in Section 3.5 an overview of the major flavour constraints on this kind of proposals. In particular, we find that adding heavier Leptoquarks to the model provides a possible solution to tensions in different low-energy observables. Another possibility is to reduce the couplings absolute value while keeping their ratio. This still provides an excess in $\mu\nu jj$ over $\mu\mu jj$, but at the price of loosing non-resonant features in the kinematic distributions.

## 3.4 BACK TO THE CMS RESULT

In previous sections we have defined our model and a region in its parameter space which we have argued to favor an excess in the final state $\mu\nu jj$ over $\mu\mu jj$. We have justified that this region in parameter space would also hide a peak in kinematic distributions because of single- and non-resonant processes involved in the production of the mentioned final states. Along this section we probe this model through a given benchmark point, Eq. 47, and test the validity of the reasoning. We first describe our simulations and then present the results, which compare our simulations with the CMS available results as reported in Section 3.1.

### 3.4.1 *Simulation*

To reproduce an excess in one channel without getting in tension with the other, and to show that single- and non-resonant production wash out the peak in the $m_{\mu j}$ distribution, we focus on a qualitative analysis at the parton level for the benchmark point detailed in Eq. 47.

We have implemented our model using `Feynrules` [14] and loaded it into `MadGraph5_aMC@NLO` [12]. We have simulated $pp \to \mu\mu jj$ and $pp \to \mu\nu jj$ including $S^{1/3}$ non-resonant, single-resonant and pair production at $\sqrt{s} = 13$ TeV. The signal events have been generated at Leading Order matrix element and with the MSTW2008 PDF set [137], setting the renormalization and factorization scales to $M_{S^{1/3}}$ [138, 139] with the same cuts as in the CMS paper Ref. [129]. After generation, expected events in each signal region are obtained assuming an efficiency of 30%. This estimation comes from the reported acceptance times efficiencies in the CMS paper Ref. [129], taking into account that detector acceptances for pair production are close to unity. Regarding NLO QCD effects we have considered the available UFO model in Ref. [119]. However, at the current level of development this model sets a maximum of one NP vertex for NLO processes and therefore it cannot generate several of the required relevant diagrams. Despite this, we have been able to compute pair production and $qg \to S^{1/3}\ell$ NLO corrections, reproducing the results of Ref. [119] which yield $K \approx 1.56$ for pair-production and $K \approx 1.38/1.3/1.65$ for $b/c/sg \to S^{1/3}j\ell$, respectively. Therefore, and taking the approximate agreement of these K-factors into account, we estimate an overall K-factor of $K = 1.5$ that includes non-resonant effects.

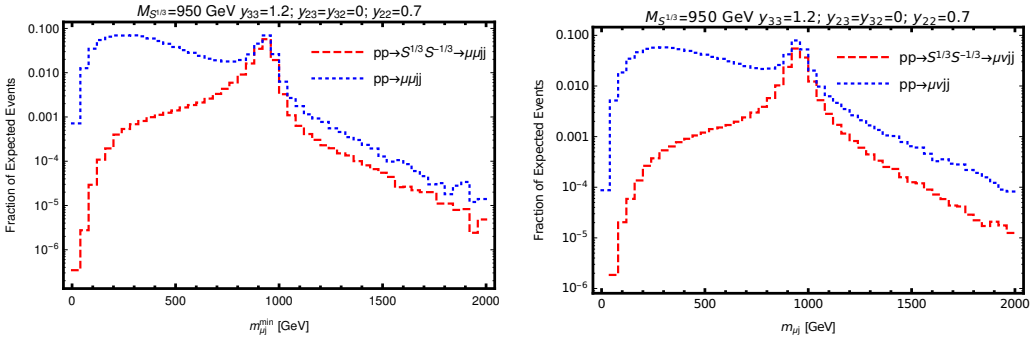
Simulations were computed in a 5-flavour scheme. However, we also performed the simulation in a 4-flavour scheme and verified that results remain qualitatively unchanged. Since we simulate up to parton level we do not need to perform a matching procedure to avoid double counting.

A more sophisticated study should include NLO generation, showering, hadronization and detector simulation. The wanted features of the model are already present at the qualitative level and are maintained as long as $r \sim 1-3$, the $y_{22} \sim y_{33} \sim \mathcal{O}(1)$ and $M_{S^{1/3}} \approx 1$ TeV. Therefore, a reasonable modification in the efficiencies could be compensated by modifying the $y_{ij}$ coefficients and/or the mass within the aforementioned values and the results presented in Subsection 3.4.2 would hold. In particular, the region around the benchmark point in Fig. 11 allows for considerable more enhancement of $\mu\nu jj$ over $\mu\mu jj$. To verify the above, we have also considered efficiencies ranging from 0.1 to 0.4, including cases with larger values for the $\mu\mu jj$ final state, and we have been able to qualitatively reproduce the same results as with the benchmark point by varying only $y_{22}$ and $y_{33}$ from 0.5 to 2.0. In any case, should the studied ex-

cess become more significant it would be essential to do a more sophisticated analysis including full signal region correlations.

### 3.4.2 *CMS Analysis*

With the simulated samples one can study the difference between considering only pair production and taking into account both single- and non-resonant behavior. This difference is made clear in the kinematic distributions, such as the $m_{\mu j}^{min}$ ($m_{\mu j}$) for the $\mu\mu jj$ ($\mu\nu jj$) final state, as can be seen in Fig. 12. It is seen in both final states that the pair production peak is washed-out by the single- and non-resonant effects. The single- and non-resonant Feynman diagrams discussed in Section 3.2 provide the kinematics to yield events with $m_{\mu j}^{min}$ and $m_{\mu j}$ considerably different to the mass of the Leptoquark involved in the process.




(a) (b)

Figure 12: $m_{\mu j}$ kinematic distribution at the a) $\mu\mu jj$ and b) $\mu\nu jj$ final states for NP-only diagrams. The red (dashed) line corresponds to pair production diagrams only, whereas the blue (dotted) line includes in addition single- and non-resonant production.

In this benchmark point, the event yield in each signal region for each final state $\mu\mu jj$ and $\mu\nu jj$ can be compared to data reported in [129]. Because of the qualitative discussion in Section 3.2 we expect NP to provide a larger contribution to the $\mu\nu jj$ than to the $\mu\mu jj$ final state.

The SM backgrounds have been taken to be those reported in Ref. [129] and they have been combined with our simulated samples in order to have a comparison as seen in Fig. 13. The top row in the figure shows both channels separated while the bottom row compares both channels together, both in absolute value and in its deviation for their respective SM background. As it can be seen, the $\mu\nu jj$ final state has both a larger absolute NP event yield, Fig. 13c, and a larger relative deviation from background than $\mu\mu jj$, Fig. 13d.

The largest deviation in the $\mu\nu jj$ final state reported in [129], located at the $M_{LQ} = 950$ GeV bin, is reduced from $2.25\sigma$ to $0.75\sigma$ in this NP benchmark point.

Therefore, in addition to the wash-out of the peak in the kinematic distributions, this shows quantitatively that the NP model can provide an explanation of the moderate excess in the $\mu\nu jj$ final state while keeping without considerable change the events in the $\mu\mu jj$ final state.

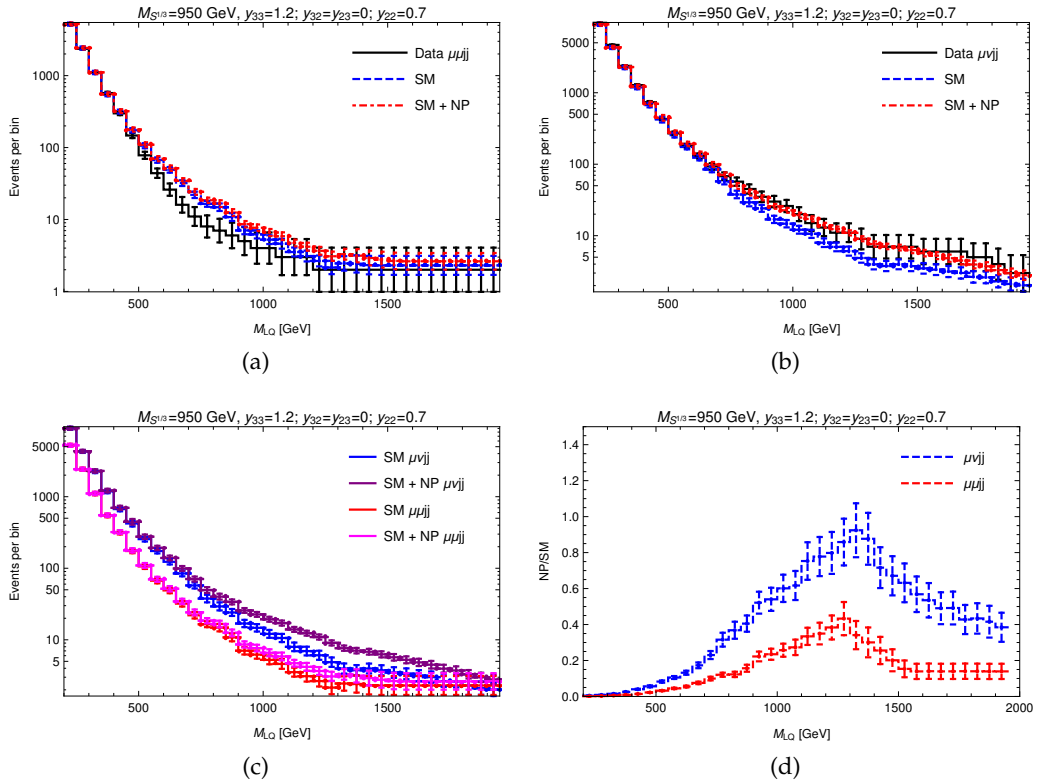



Figure 13: Top row: Comparison between data, SM and SM+NP events for a) μμjj and b) μνjj final states. Bottom row: Comparison between μμjj and μνjj NP event yields in a) absolute value b) relative deviation from SM

## 3.5 FLAVOR CONSTRAINTS

Along this Section we examine constraints coming from low-energy flavour physics that could enter into conflict with the large couplings required in the model presented so far. Indeed, such large couplings of $\mathcal{O}(1)$ would be ruled out by flavour constraints. This could be avoided if other heavier particles are added to the model to cancel low-energy bounds while slightly affecting collider observables. As a matter of fact, since the low-energy contributions are in general inversely proportional to the squared mass whereas contributions to collider physics observables have an exponential suppression in the mass due to the energy availability in the pdfs, this kind of cancellation is in general possible.

Since at this stage a new particle would be added ad-hoc to cancel the low-energy constraints, and in order to emphasize that we are presenting merely a proof of concept for the second generation Leptoquark searches, we explicitly present this alteration of the original model only in this Section. We aim to show that there exists a relatively vast region of parameter space that allows us to have the required cancellations at low-energy while still showing the non-resonant observed excess in μνjj.

We take as a departure point the model presented in Section 3.2. The running of the couplings from the TeV scale to the low-energy due to the renormalization group equations can be computed as in Ref. [140–142]. However, we expect it to be a small contribution and hence we assume it is sub-leading as is usually done [141, 142]. In any case, in the following paragraphs we show that the cancellations that take place

due to adding a new particle leave a freedom of choice that could cancel this kind of contributions as well.

The most severe constraint to the model from low-energy observables comes from the sensitive decay $B \to K^{(*)}\nu\nu$ [46, 48] which was constrained to be [143]

$$\begin{aligned} R_{\nu\nu}^{(*)} &= \frac{BR(B \to K^{(*)}\nu\nu)}{BR(B \to K^{(*)}\nu\nu)^{SM}} \\ R_{\nu\nu} &< 3.9 \\ R_{\nu\nu}^{*} &< 2.7 \end{aligned} \tag{48}$$

at 90% confidence level. The relevant Leptoquark effects can be parametrized as

$$\begin{aligned} R_{\nu\nu}^{*} &= \frac{\sum_{i,j} |\delta_{ij} C_L^{SM} + \delta C_L^{ij}|^2}{3|C_L^{SM}|^2} \\ C_L^{SM} &= -6.38(6) \\ \delta C_L^{ij} &= \frac{\pi v^2}{2\alpha_{em} V_{tb} V_{ts}^*} \frac{y_{3j} y_{2i}^*}{M_{S^{1/3}}^2}, \end{aligned} \tag{49}$$

where the sum is over the neutrinos flavour. Considering our ansatz and the central values reported in Ref. [144] for every quantity, Eq. 48 yields

$$\frac{|y_{33} y_{22}^*|}{M_{S^{1/3}}^2} < 0.045 \left(\frac{1}{\text{TeV}}\right)^2. \tag{50}$$

where we used the central value for $V_{ts}$ coming from indirect measurements assuming CKM unitarity [144]. However, for the sake of completeness, we also show in the following analysis the limits coming from direct bounds on $V_{ts}$ due to measurements of b-jet fractions in $t \to Wj$ which is $V_{ts} < 0.21$ at 95% C.L. [145, 146]; and does not assume CKM unitarity.

Since our original model ansatz in Section 3.2 would in general not satisfy bounds in Eq. 50, we add a second heavier $S_1^{'}$ with the same ansatz but different couplings $y_{22,33}^{'}$ and all others $y_{ij}^{\prime} = 0$. This converts the bounds in Eq. 50 into

$$|\frac{y_{33} y_{22}^*}{M_{S^{1/3}}^2} + \frac{y_{33}^{'} y_{22}^{'*}}{M_{S^{1/3}}^{\prime 2}}| < 0.045 \left(\frac{1}{\text{TeV}}\right)^2. \tag{51}$$

We show below how a correct relative sign assignment to the couplings allows us to obtain a cancellation to the flavour contribution and satisfy Eq. 51 while having non-resonant effects at the collider observables $\mu\mu jj$ and $\mu\nu jj$.

Regarding other flavour constraints in this new model with a second heavier Leptoquark, charged currents $b \to c\ell\nu$ effects must be considered. LFU tests give hints on Leptoquarks that couple to $\tau$-leptons while constraining couplings to $e$, $\mu$ leptons. The chosen ansatz does not produce scalar and tensor operators in the effective theory due to the absence of right-handed couplings and so the Leptoquark contribution only rescales the SMEFT [35, 46, 48, 50]. For $B \to D_s^{(*)}\ell\nu$, the Leptoquark leading contribution is proportional to

$$y_{3\ell}(y_{3\ell}^* + \frac{V_{cs}}{V_{cb}} y_{2\ell}^*) \tag{52}$$

It is easy to see that this vanishes for $\ell = e, \mu$ but not for $\tau$. Considering $R_{D^*}$, the Leptoquark effects at leading order are

$$\begin{aligned} \frac{R_D}{R_D^{SM}} &= \frac{R_{D^*}}{R_{D^*}^{SM}} = 1.237 \pm 0.053 \approx 1 + 2g_{V_L} \\ g_{V_L} &= \frac{v^2}{4}\left(\frac{|y_{33}|^2}{M_{S^{1/3}}^2} + \frac{|y'_{33}|^2}{M'^2_{S^{1/3}}}\right). \end{aligned} \tag{53}$$

Therefore $R_{D^*}$ implies

$$2.466 < \sqrt{\left(|y_{33}|\frac{\text{TeV}}{M_{S^{1/3}}}\right)^2 + \left(|y'_{33}|\frac{\text{TeV}}{M'_{S^{1/3}}}\right)^2} < 3.096 \tag{54}$$

at the $1\sigma$ level.

Charged current effects are also seen in meson decays such as $B_c \to \ell\nu$ and $D_s \to \ell\nu$ [35, 50]. The only relevant decays in the benchmark point (due to $y_{23}^{(\prime)} = y_{32}^{(\prime)} = 0$) at leading order in $v/M_{S^{1/3}}^{(\prime)}$ are

$$\begin{aligned} \Gamma_{B_c\to\tau\nu} &= \frac{G_F^2}{8\pi} f_{B_c}^2 m_{B_c}^3 (1 - \frac{m_\tau^2}{m_{B_c}^2})^2 \frac{m_\tau^2}{m_{B_c}^2} |V_{cb}|^2 (1 + \frac{v^2}{2M_{S^{1/3}}^2}|y_{33}|^2 + \frac{v^2}{2M'^2_{S^{1/3}}}|y'_{33}|^2) \\ \Gamma_{D_s\to\mu\nu} &= \frac{G_F^2}{8\pi} f_{D_s}^2 m_{D_s}^3 (1 - \frac{m_\mu^2}{m_{D_s}^2})^2 \frac{m_\mu^2}{m_{D_s}^2} |V_{cs}|^2 (1 + \frac{v^2}{2M_{S^{1/3}}^2}|y_{22}|^2 + \frac{v^2}{2M'^2_{S^{1/3}}}|y'_{22}|^2) \end{aligned}$$

Constraints from these decays require form factors computed by lattice QCD [144]. These expressions show that Leptoquark effect in these decays produce a rescaling of $V_{cb}$ and $V_{cs}$. Assuming that this scaling could be detected within the measured uncertainty in the CKM elements, the more relevant constraint would come from $V_{cs}$, which is measured with a percent-level precision [144]. Therefore,

$$|V_{cs}| \to |V_{cs}|(1 + \frac{v^2}{4M_{S^{1/3}}^2}|y_{22}|^2 + \frac{v^2}{4M'^2_{S^{1/3}}}|y'_{22}|^2) \tag{55}$$

implies

$$\left(|y_{22}|\frac{\text{TeV}}{M_{S^{1/3}}}\right)^2 + \left(|y'_{22}|\frac{\text{TeV}}{M'_{S^{1/3}}}\right)^2 < 1.13 \tag{56}$$

at the $1\sigma$ level.

Having expressed quantitatively the constraints coming from low-energy precision physics, we proceed to find possible solutions that provide non-resonant effects in collider observables while avoiding these flavour constraints. In order to explicitly construct a model, we choose a new benchmark point BP$'$ where one Leptoquark has mass $M_{S^{1/3}} = 950$ GeV and the second Leptoquark has mass $M'_{S^{1/3}} = 1500$ GeV. The only non-zero couplings are $(y_{22}, y_{33}) = (0.7, 1.2)$ and $(y'_{22}, y'_{33}) = (0.7, 3)$, which yields an exact cancellation in Eq. 51 while still having couplings below the perturbative limit. Performing the same simulation as in Section 3.4, we obtain the results shown in Fig. 14. From this Figure we see that this model with two Leptoquarks again reproduces the sought phenomenology regarding the CMS paper Ref. [129]. As a matter of fact, the presence of the heavy Leptoquark is hardly seen in the collider phenomenology. The excess in the $M_{LQ} = 950$ GeV bin in $\mu\nu jj$ is reduced in this case

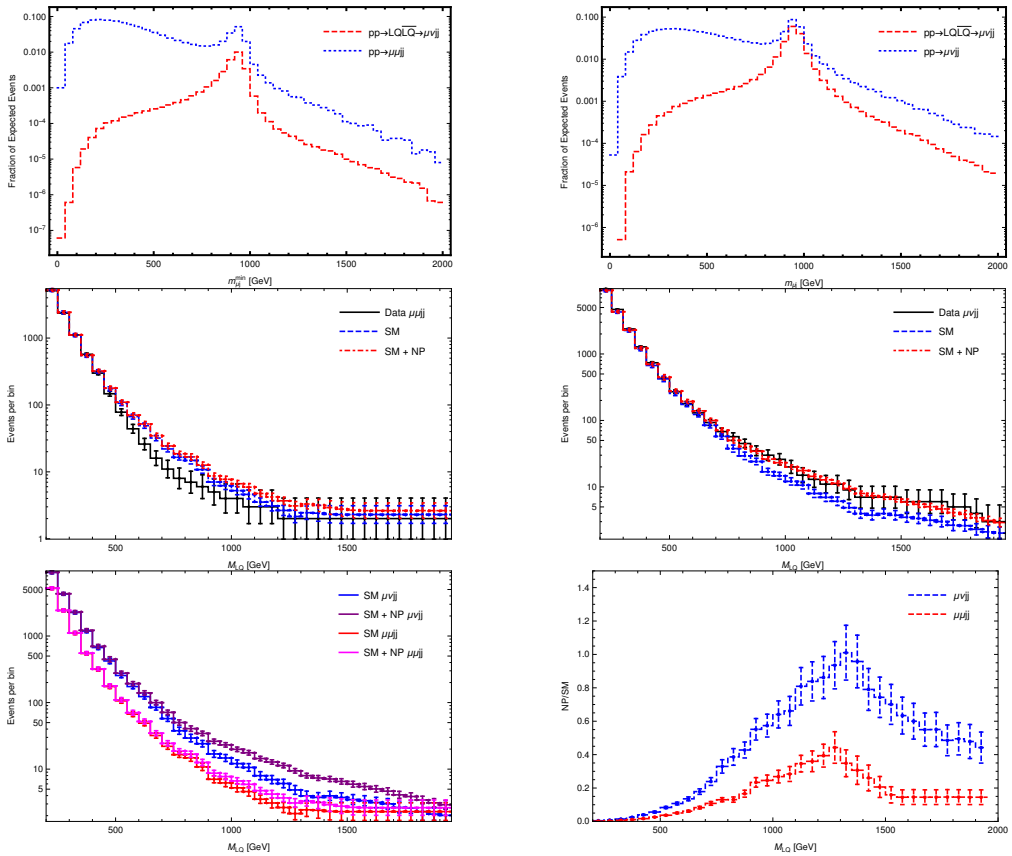



Figure 14: Collider phenomenology of two $S_1$ in the benchmark point $(M_{S^{1/3}}, y_{22}, y_{33}) = (950\ \text{GeV}, 0.7, 1.2)$ and $(M'_{S^{1/3}}, y'_{22}, y'_{33}) = (1500\ \text{GeV}, 0.7, 3)$. This model avoids the discussed flavour constraints, retaining certain freedom in the parameters of the heavier Leptoquark as shown in Figs. 15 and 16.

from 2.25 to 0.65 standard deviations, in contrast to $0.75\sigma$ for the case of only one Leptoquark.

Since the main collider phenomenology is produced by the 950 GeV Leptoquark, we can examine up to which extent the parameters of the heavier Leptoquark are determined by the flavour constraints. In Fig. 15 we study the parameter space for the heavier Leptoquark that satisfies the $K^{(*)}\nu\nu$ constraint. In Fig. 16 we do the same for constraints coming from $R_D^{(*)}$ and $V_{cs}$.

Summarizing, we have verified that it is possible to have non-resonant effects in $\mu\mu jj$ and $\mu\nu jj$, with an excess in the latter, while avoiding the discussed flavour constraints. Of course our model has too many arbitrary features, but it could be useful to study up to what extent these features could be obtained from a more complete theory.

## 3.6 DISCUSSION

The results obtained in Section 3.4 reflect the differences qualitatively discussed in Section 3.1 between the two channels studied in second generation Leptoquark searches when the proposed NP has a multigeneration non-resonant phenomenology.

The chosen benchmark point includes the simplistic assumptions made in Section 3.2. If one seeks a flavour hierarchical model, this may require non negligible $y_{32}$ and $y_{23}$. If $V_{CKM} \approx 1$, $y_{23}$ opens the $c\tau$ channel while $y_{32}$ opens the $t\mu$

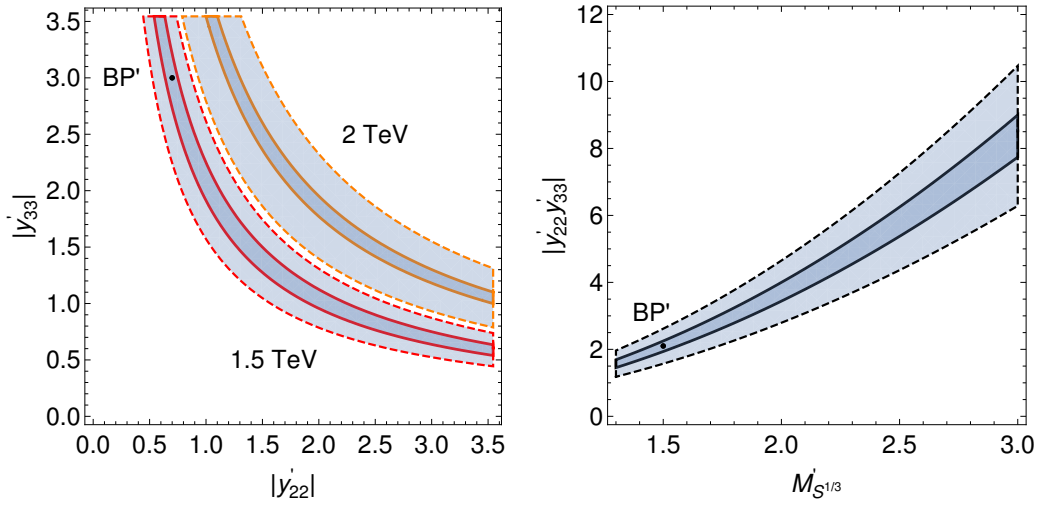



Figure 15: Allowed region in parameter space of the heavier Leptoquark where the $K^{(*)}\nu\nu$ constraint detailed in Eq. 51 is satisfied once the lighter Leptoquark is fixed to $(M_{S^{1/3}}, y_{22}, y_{33}) = (950\ \text{GeV}, 0.7, 1.2)$. In the left panel we consider two possible masses for the heavier Leptoquark. The solid lines represent the 90% C.L. allowed using the central value for $V_{ts}$ assuming CKM unitarity, whereas the dashed lines correspond to the direct limit on $V_{ts}$ as detailed in text.

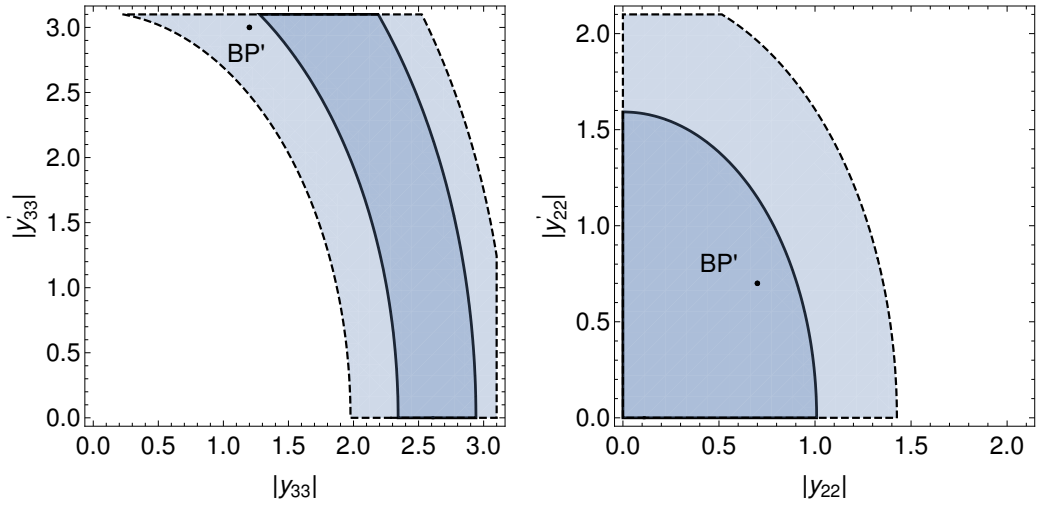



Figure 16: *Left:* Allowed region in parameter space of the heavier Leptoquark where the $R_D^{(*)}$ constraint detailed in Eq. 54 is satisfied for $M'_{S^{1/3}} = 1.5$ TeV and $(M_{S^{1/3}}, y_{22}, y_{33}) = (950\ \text{GeV}, 0.7, 1.2)$. *Right:* idem for constraints coming from $V_{cs}$ precision, detailed in Eq. 56. The solid and dashed lines represent the $1\sigma$ and $2\sigma$ levels, respectively. In both cases the mark corresponds to the selected benchmark point.

channel. This enlarges the Leptoquark width, diminishing $BR(j\mu)$ while maintaining $\sum_i BR(j\nu_i) = 1/2$ because of the neutrino mixing matrix. As the final states do not target top quarks nor tau leptons, this increases even more the difference between $\mu\mu jj$ and $\mu\nu jj$ final states in pair production. Single- and non-resonant production remains relatively unchanged if one does not alter the values of $y_{22}$ and $y_{33}$.

In Section 3.2 we decided to work with the $Q = 1/3$ Leptoquark, although also a $Q = 2/3$ Leptoquark could yield the same final states. The reason for this decision is clear after the discussion in Section 3.2. A Leptoquark with charge $Q = 2/3$ would require more up-type instead of down-type quarks in the initial state to enhance the $\mu\nu jj$ final state, since this Leptoquark connects the neutrinos with $Q = 2/3$ quarks. This is unfeasible as we neglect first generation couplings. Even more, the third generation decay would be to $\tau\bar{b}$ instead of $\nu\bar{b}$. In this case, acceptance analyses should be performed after the $\tau$ decay to investigate in which specific cases one could obtain an enhancement of $\mu\nu jj$ over $\mu\mu jj$. Therefore, the reasonable decision is to assume a Leptoquark with electric charge $Q = 1/3$ in which case all the wanted features come out naturally.

Along this Chapter we have focused on the excess reported by the CMS paper Ref. [129] in the $\mu\nu jj$ final state. However, the same paper reports a deficit in the $\mu\mu jj$ final state at lower signal regions $M_{LQ} \approx 600\,\text{GeV} \sim 800\,\text{GeV}$. If one tries to give a NP interpretation of this deficit, single- and non-resonant Leptoquark production are convenient because they can interfere with the SM. This interference is larger for smaller values in $M_{LQ}$. By computing the interference between NP and the SM main background, Z+jets, with `MadGraph5_aMC@NLO` [12] we have found that in the benchmark point detailed in Eq. 47 the interference is always destructive in $\mu\mu jj$. However, the strength of this interference is too small to account for the deviation. As we added a second Leptoquark to avoid flavour constraints, we have also investigated if another Leptoquark of mass $\sim 600$ GeV could be producing such an effect. We find in general that the interference in the $\mu\mu jj$ final state is destructive, but is not possible to reach the observed strength in the data while keeping the model from being ruled out by other direct Leptoquark searches. It is interesting to notice that, on the other hand, the $\mu\nu jj$ final state interferes with the SM background $W$+jets, but its sign can be adjusted by the relative sign between the PMNS component $U_{22}$ and the CKM matrix. As a curiosity, if one compares $S_3^{1/3}$ and $S_1^{1/3}$ as separate candidates with the same coupling constants, the interference in $\mu\nu jj$ is of opposite sign. Thus, $\mu\nu jj$ production could provide a handle to distinguish between different Leptoquark species.

As for the anomalous event reported in the $\mu\mu jj$ final state, we have not performed an exhaustive analysis as this information was added to the CMS paper Ref. [129] after we submitted Ref. [130]. However, we would like to discuss single- and non-resonant effects on $S_T^{\mu\mu jj}$. As seen in Fig.17, single- and non-resonant production yield smaller $S_T^{\mu\mu jj}$ values than pair production. This can be understood when thinking of $S_T^{\mu\mu jj}$ as a measure of the activity of the event, and as such centered around $2M_{S^{1/3}}$ for pair production and $M_{S^{1/3}}$ for single-resonant production. To obtain an event where the selected two $\mu j$ pairs have a different invariant mass, which would be characteristic of single- and non-resonant production, it would need to originate from a heavier $S_1$ than the one needed to accommodate the excess in $\mu\nu jj$. Even more, to yield an event with $S_T^{\mu\mu jj} > 3000\text{GeV}$, the Leptoquark would need to be even heavier than the $S_1'$ introduced in Section 3.5. However, collider effects for such a heavy Leptoquark are so suppressed than an observed event would be extremely rare. In light of this, adding single- and non-resonant production alone does not seem to be able to accommodate such a large $S_T^{\mu\mu jj}$. However, this argument highlights the fact that because the cuts on $S_T$ for each $M_{LQ}$ are optimized for pair production, they leave out many single- and non-resonant events which would appear in lower $M_{LQ}$ signal regions as events with a $m_{\mu j}^{(\text{min})}$ different from $M_{S^{1/3}}$.

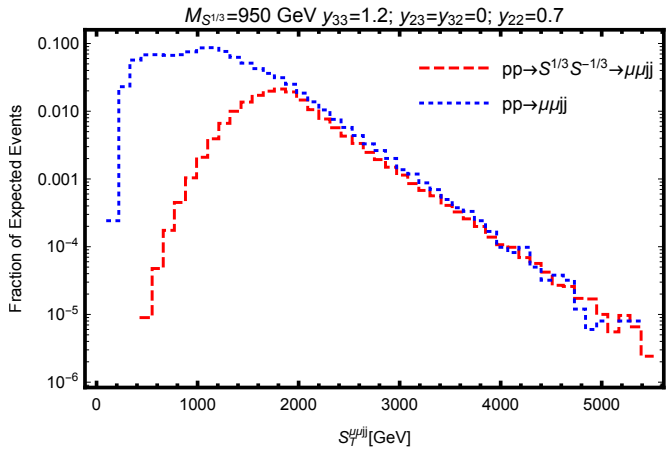



Figure 17: $S_T^{\mu\mu jj}$ kinematic distribution for NP-only diagrams. The red (dashed) line corresponds to pair production diagrams only, whereas the blue (dotted) line includes in addition single- and non-resonant production.

## 3.7 CONCLUSIONS

Along this Chapter we have considered non-resonant, single-resonant and pair production effects from a Leptoquark with diagonal couplings to second and third generation in the final states $\mu\mu jj$ and $\mu\nu jj$. We have shown that a non-resonant excess in the $\mu\nu jj$ final state and a fainter excess in the $\mu\mu jj$ final state is a pattern of this NP if the couplings are large enough and couple more to the third than to the second generation.

We have presented a simple Leptoquark model with $\mathcal{O}(1)$ couplings to second and third generation and a mass of $\mathcal{O}(1\text{ TeV})$. On one hand, the strength of these couplings assures that single- and non-resonant effect are important and any kinematic resonant effect in the final states is spread out. On the other hand, a larger coupling to third generation favors the $\mu\nu jj$ final state because of two reasons. First, as long as there is no b-jet veto, since the $\nu$ is not flavour-tagged, a produced Leptoquark can decay to $j\nu$ where $j$ can be a $b$. Second, single- and non-resonant production requires quarks in the initial state, and there is a larger abundance of the down-type quarks that produce the $\mu\nu jj$ final state than their same generation up-type quarks, whose corresponding diagrams would yield $\mu\mu jj$; with $\mu\nu jj$ having also $gc$ initial states. We have performed a simple simulation including the NP indicated by this model and compared our results to the CMS paper Ref. [129], where a non-resonant excess in the $\mu\nu jj$ has been reported. We have found that this simple Leptoquark model can easily accommodate the data better than SM alone and provide the features previously described. In particular, if we take the bin reported as the largest excess in $\mu\nu jj$ in CMS paper Ref. [129], then our model reduces the significance from $2.25\sigma$ to $0.75\sigma$ while the $\mu\mu jj$ final state is barely affected.

This Chapter indicates that an excess of this kind, if produced by NP, would not be properly distinguished by current second generation Leptoquark searches at the LHC. The analysis in this Chapter points out that to observe a NP signal of this kind, experiments should test the effect of b-tagging the jets. In that case the $\mu\nu jj$ sample will enhance its excess for the b-tagged events, whereas the excess should diminish in the b-tag veto sample. Moreover, if the statistic is large enough then pairing the b-jet

with the $\nu$ and the light-jet with the $\mu$ in the $\mu\nu jj$ sample could enhance a possible peak depending on the absolute value of the couplings. In addition to all of this, this analysis highlights the fact that $\mu\nu jj$ has a different phenomenology than $\mu\mu jj$ and it should always be included as a final state on second generation Leptoquark searches.

We have also discussed general features of the NP model. We find that the parameter space of the model has room for variations and still reproduce the same qualitative results. We have found that the large couplings required to yield non-resonant effects would enter into conflict with low-energy observables. We have shown how adding a new heavier Leptoquark could avoid low-energy flavour problems, and present the details of a model where this is accomplished. We also discussed another feature of the CMS paper Ref. [129] that consists of a deficit in the $\mu\mu jj$ final state with respect to the SM expected background in the signal regions $M_{LQ} \approx 600$ GeV $\sim$ 800 GeV. We show that SM and NP interference has negative sign in this final state. However we cannot fulfill the observed deficit within this simple Leptoquark model while keeping the model from being ruled out by other direct searches. It would be interesting to further investigate these or others Leptoquark models to also accommodate the deficit in $\mu\mu jj$.

Summarizing, we have shown that current second generation Leptoquark searches in the $\mu\mu jj$ and $\mu\nu jj$ final states are also sensitive to diagonal couplings to third generation as long as a $b$-jet veto is not applied. We have presented a simple model that reproduces an excess recently found in CMS and that was disregarded because only second generation couplings were taken into account. The results presented in this Chapter suggest that Leptoquark searches at the LHC should take into account single- and non-resonant effects and multigenerational couplings.

# 4

# Z′ HIDING IN TTH AND 4-TOP

In this Chapter we propose a purely phenomenological BSM model where a Z′ spin 1 boson with mass below 1 TeV is responsible for reported discrepancies in same sign dilepton and multilepton final states with high b-jet multiplicity. We choose this model with a particular coupling structure that is able to provide both charge asymmetry and high b-jet multiplicity. These two features are the key ingredients to accommodate the different experimental signatures. We also study how such a model would appear in a global search and propose a new kinematical variable to improve the sensitivity.

This Chapter is organized as follows. We set the stage by detailing the aforementioned discrepancies in same sign dilepton and multilepton final states with high b-jet multiplicity in Section 4.1. We present our phenomenological model aimed to explain these discrepancies in Section 4.2. We then study its experimental imprints in Section 4.3, including estimating the region of parameter space that best fits the $t\bar{t}H$ and 4-top ATLAS data, and possible constraints from other observables on this parameter space. In Section 4.4 we propose a global search strategy to discriminate signal from background. Finally, we present our conclusions in Section 4.5.

The basis of this Chapter consists of work done in collaboration with Ezequiel Alvarez, Aurelio Juste and Tamara Vazquez Schroeder which resulted in Ref. [147]. I took part of every part of the work involved except on the likelihood fits to the $t\bar{t}H$ experimental data and the expected significance and exclusion limits calculations of the Z′ using an Asimov dataset [148], all of which were performed with the help of the `HistFitter` package [124].

## 4.1 RECENT DISCREPANCIES IN $t\bar{t}W^{\pm}$, $t\bar{t}H$ AND 4-TOP SEARCHES

As detailed in Subsection 1.1.1, same sign dilepton and multilepton final states with high b-jet multiplicity are a very interesting final state to explore rare SM processes because they present a good balance between low SM backgrounds and high signal yields. By targeting these final states, the LHC program has made impressive progress in the measurement of the cross-sections of $t\bar{t}W^{\pm}$[149, 150], $t\bar{t}H$ [151, 152] and 4-top ($t\bar{t}t\bar{t}$) [25, 26] production. Intriguingly, there has been a persistent discrepancy in the $t\bar{t}W^{\pm}$ normalization, as the observed yields are consistently larger than those expected from state-of-art theoretical predictions. This discrepancy has been found both in dedicated $t\bar{t}W^{\pm}$ cross-section measurements [149, 150], and in several searches that consider the $t\bar{t}W^{\pm}$ process as a background [25, 26, 151, 152], as summarised in Table 3. We see that the ATLAS and CMS $t\bar{t}H$ measurements [151, 152] analyse a larger dataset than the $t\bar{t}W^{\pm}$ cross-section measurements [149, 150]. If we consider the same reference cross-section for all searches, using for example the reported cross-section of Ref. [153], we see that the $t\bar{t}H$ measurements show a larger tension than the $t\bar{t}W^{\pm}$ measurements. Because of this two facts, we focus on the $t\bar{t}H$ searches. Despite having similar strategies to suppress non-prompt lepton background, there are several differences in the $t\bar{t}H$ analysis approach between ATLAS and CMS, such as different fake estimation techniques (simultaneous profile likelihood template fit in ATLAS vs misidentification probability method in CMS) and different analysis strategies (jet

multiplicity, total lepton charge, and Boosted Decision Trees (BDT) categorisation in ATLAS vs Deep Neural Networks in CMS). A summary of the event selection used in the ATLAS and CMS $t\bar{t}H$ analyses can be found in Table 4. We consider only the two same sign dilepton and trilepton channels where hadronically decaying $\tau$s are vetoed.

We choose the ATLAS $t\bar{t}H$ measurement for the re-interpretation in this Chapter since it provides complete information on the multilepton discrepancies observed as a function of the total lepton charge and the b-jet multiplicity of the event. A recent CMS search for new physics within the effective field theory (EFT) framework [154] using 41.5 $fb^{-1}$ of integrated luminosity exploits a similar categorisation. However, since there is no information on the two same sign dilepton selection with exactly 1 b-jet and given the lower data statistics, we do not include the CMS EFT results in this Chapter.

In addition, a 4-top search by the ATLAS Collaboration in the same final states has measured a 4-top cross-section a factor of two higher than the SM prediction [25]. A summary of the signal region event selection used in the ATLAS and CMS 4-top analyses can be found in Table 5.

It should be pointed out that these measurements are individually consistent with the SM and thus the observed discrepancies could be merely the result of statistical fluctuations, and/or unaccounted experimental or theoretical uncertainties. However, when combined, they paint an interesting picture that is worth exploring, as it may open the door to new exciting discoveries.

| Search | $\mathcal{L}$ [$fb^{-1}$] | $\sigma_{ref}$ [pb] | $\mu$ | $\mu_{YR_4}$ |
|---|---|---|---|---|
| $t\bar{t}W^{\pm}$ ATLAS [149] | 36.1 | $0.60 \pm 0.07$ | $1.44 \pm 0.32$ | $1.44 \pm 0.32$ |
| $t\bar{t}W^{\pm}$ CMS [150] | 35.9 | $0.628 \pm 0.082$ | $1.23^{+0.30}_{-0.28}$ | $1.29^{+0.31}_{-0.29}$ |
| $t\bar{t}H$ ATLAS [151] | 80 | $0.727 \pm 0.092$ | $1.39^{+0.17}_{-0.16}$ | $1.68^{+0.21}_{-0.19}$ |
| $t\bar{t}H$ CMS [152] | 137 | 0.650 | $1.43 \pm 0.21$ | $1.55 \pm 0.23$ |
| 4-top ATLAS [25] | 139 | 0.601 | $1.6 \pm 0.3$ | $1.6 \pm 0.3$ |
| 4-top CMS [26] | 137 | 0.610 | $1.3 \pm 0.2$ | $1.3 \pm 0.2$ |

Table 3: This table lists the $t\bar{t}W^{\pm}$ reference cross-sections and the corresponding signal strengths for different searches. The last column is the signal strength corresponding to the reference cross-section listed in Ref. [153], 600.8 fb. The reference cross-section was not listed in Ref. [26] and so it was taken from Ref. [155].

## 4.2 THE MODEL

We aim to construct a model that can account for several slight excesses in same sign dilepton and multilepton plus b-jets final states at the LHC. More precisely, we require the model to be capable of producing more positive than negative leptons to better fit the imbalance suggested by the results in Ref. [151]. We motivate the model from a phenomenological point of view and then examine its theoretical strengths.

Since the excess reported in Ref. [151] is charge imbalanced, with a larger excess in the channels with positive total leptonic charge than in the channels with negative total leptonic charge, the model needs to capture the excess in positive charge present in the colliding protons. Thus the BSM should couple to u quarks while being safe

| $t\bar{t}H$ | ATLAS | | CMS | |
|---|---|---|---|---|
| | 2LSS | 3L | 2LSS | 3L |
| Total lepton charge | $\pm2$ | $\pm1$ | $\pm2$ | $\pm1$ |
| Lepton $p_T$ [GeV] | 20/20 | 15/15/10 | 25/15 | 25/15/10 |
| Number of jets | $\geqslant 2$ | | $\geqslant 3$ | $\geqslant 2$ |
| Number of b-jets | $\geqslant 1$ (70% eff.) | | $\geqslant 1$ (70% eff.) / $\geqslant 2$ (84% eff.) | |
| $\lvert m_{\ell\ell}\rvert$ (2LSS) or $\lvert m_{OSSF}\rvert$ (3L) [GeV] | $> 12$ | | | |
| $\lvert m_{e^\pm e^\pm} - m_Z\rvert$ (2LSS) or $\lvert m_{OSSF} - m_Z\rvert$ (3L) [GeV] | - | $> 10$ | | |
| Other | - | $\lvert m_{\ell\ell\ell} - m_Z\rvert > 10$ GeV | Missing transverse momentum cuts | |

Table 4: Comparison of event selections between the ATLAS [151] and CMS [152] $t\bar{t}H$ analyses.

| 4-top | ATLAS | | CMS | |
|---|---|---|---|---|
| | 2LSS | $\geqslant$ 3L | 2LSS | $\geqslant$ 3L |
| Total lepton charge | $\pm2$ | - | $\pm2$ | - |
| Lepton $p_T$ [GeV] | 28 (all $\ell$) | | 25/20 | 25/20/20(/20) |
| Number of jets and b-jets | $\geqslant$ 6j $\geqslant$ 2bj (77% eff.) | | $\geqslant$ 6j $\geqslant$ 2bj / 5j $\geqslant$ 3bj (55-70% eff.) | $\geqslant$ 5j $\geqslant$ 2bj / 4j $\geqslant$ 3bj (55-70% eff.) |
| $H_T$ [GeV] | $> 500$ | | $> 300$ | |
| $\lvert m_{e^\pm e^\pm}\rvert$ (2LSS) or $\lvert m_{OSSF}\rvert$ (3L) [GeV] | $> 15$ | - | $> 12$ | |
| $\lvert m_{e^\pm e^\pm} - m_Z\rvert$ (2LSS) or $\lvert m_{OSSF} - m_Z\rvert$ (3L) [GeV] | $> 10$ | | - | $> 15$ |
| Other | - | | Missing transverse momentum cuts | |

Table 5: Comparison of event selections between the ATLAS [25] and CMS [26] 4-top analyses. $H_T$ is the scalar $p_T$ sum of jets, leptons and b-jets.

to low-energy physics observables. Should the new particle coupling to u quarks be charged then it should also couple to bottom quarks (a $W'$ boson) or to leptons (a Leptoquark). If the new particle were a $W'$, then it is difficult to obtain a multilepton final state with leptonic charge asymmetry and high b-jet multiplicity without including large first generation $g_{ud}$ couplings. Were the new particle a Leptoquark then, being charged under $SU(3)_C$, the bounds on its mass from pair production would

make it more difficult to reproduce the observed excesses. We are then left with neutral particles coupling to up-type quarks (u, c and t) and with non-negligible Flavour Changing Neutral Currents (FCNC) to account for the deviations. Among the three usual spins 0, 1 or 2, we find suitable to study spin 1 since, if colour neutral, its gluon fusion production is protected through the Landau-Yang theorem [156–158], which also guarantees that the restrictive bounds in di-photon do not apply [159–163]. Examples of studies of a FCNC spin 0 scalar boson phenomenology at the LHC can be seen in Refs. [164, 165].

We consider a FCNC $Z'$, a colour neutral spin 1 particle of the kind detailed in Subsection 1.2.2. There, we wrote its most general interaction Lagrangian with couplings restricted to up-type quarks [5]:

$$\mathcal{L}_{int} \supseteq Z'_\mu \left( \sum_{i,j=u,t,c} \bar{u}_i \gamma^\mu (g^L_{ij} P_L + g^R_{ij} P_R) u_j \right). \tag{57}$$

We refer to it as a FCNC $Z'$ due to its with non-diagonal couplings. We are omitting the necessary kinetic Lagrangian and we neglect any possible Z-$Z'$ mixing because of the restrictive bounds imposed by LEP [61].

$$\mathcal{L}_{int} \supseteq Z'_\mu \left( g_{ut}\, \bar{t}_R \gamma^\mu u_R + g_{ct}\, \bar{t}_R \gamma^\mu c_R + g_{tt}\, \bar{t}_R \gamma^\mu t_R \right) + \text{h.c.}, \tag{58}$$

where we have dropped the R supra-index to the couplings to simplify notation.

Effective theories similar to the one described by the interaction in Eq. 58 have been studied in different contexts, as for instance in Refs. [28, 64, 166–168]. In particular, a very similar set-up has been implemented in Ref. [169], albeit in a different mass range and assuming a hidden sector that increases the width-to-mass ratio, and in Ref. [170] where a similar model was implemented to study the forward-backward $t\bar{t}$ asymmetry at the Tevatron.

### 4.2.1 *Phenomenology*

To account for the observed data (see Subsection 4.3.1 for more details), we need to produce same electric charge dilepton (denoted 2LSS, with SS standing for same sign) and multilepton (at least three leptons, denoted 3L) final states with charge asymmetry and high b-jet multiplicity. To accomplish this, we consider two relevant $Z'$-induced processes, $tZ' + \bar{t}Z'$ (denoted $tZ'$ in the following) and $t\bar{t}Z'$, with a hierarchy between the relevant couplings to enforce a high probability of three- and four-top-quarks final states. We show two representative examples of the relevant Feynman diagrams in Fig. 18, and the relevant cross-sections and Branching Ratios in Fig. 19 as a function of $M_{Z'}$ for a given set of couplings. The cross-sections for a different set of couplings can be obtained by simple re-scaling. The Branching Ratios for a different set of couplings can be obtained by using $\Gamma(Z' \to t\bar{u})/(g_{ut})^2 = \Gamma(Z' \to t\bar{c})/(g_{ct})^2$ and re-scaling.

Other processes that could be tested in these channels (see Subsection 4.3.2) are either numerically irrelevant such as same sign top-quark pair production ($tt$, $\bar{t}\,\bar{t}$) [169], or chirality suppressed such as radiative $Z'$ production ($tZ'j$, $\bar{t}Z'j$) [169, 171]. Note that these results can be fairly specific to our model, see e.g. Ref. [172] for a different $Z'$ model where $cg \to tZ'$ is the main production channel. In particular, non-resonant effects are suppressed both by the small $Z'$ width we consider by neglecting hidden sector decays, and by our particular choice of coupling structure.

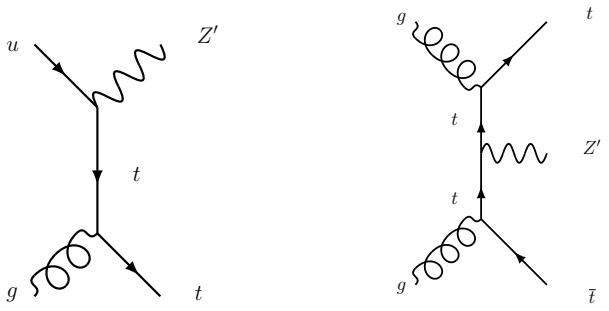
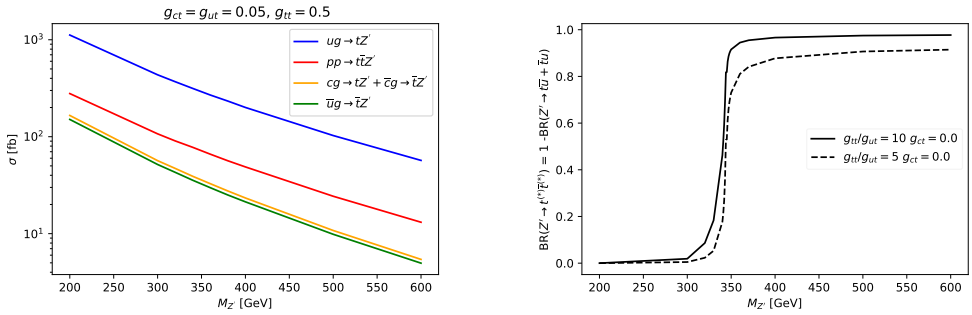



Figure 18: Representative diagrams for some of the most relevant processes: (a) $tZ'$ production and (b) $t\bar{t}Z'$ production.




(a) (b)

Figure 19: (a) Production cross-sections at $\sqrt{s} = 13$ TeV as a function of $M_{Z'}$ for the most relevant signal processes, assuming a benchmark set of couplings. (b) Branching Ratios for the dominant $Z'$ decay modes as a function of $M_{Z'}$ for two different benchmark sets of couplings.

From Fig. 18, we see that contributions to $tZ'$ are proportional to $(g_{ut})^2$ and to $(g_{ct})^2$ while $t\bar{t}Z'$ is proportional to $(g_{tt})^2$. From Fig. 19a, we see that even for $g_{tt} \gg g_{ut}, g_{ct}$, the cross-section for $tZ'$ production is larger than for $t\bar{t}Z'$, mainly due to the kinematic requirements that must be met to produce each of the final-state particles on-shell. When comparing the four possible $tZ'$ processes, the largest cross-section is $ug \to tZ'$, as expected from the model motivations. This is due to $u$ quark abundance in the proton, which ensures that the $g_{ut}$-induced processes yield a considerable charge asymmetry that is not present in the other production processes. We are interested in this charge asymmetry and how it is reflected in current experimental searches.

If we take the possible relevant decays obtained from the Lagrangian in Eq. 58 into account, and assuming they are all kinematically accessible, we see that $tZ'$ can yield the following final states: $ttj$, $\bar{t}\,\bar{t}j$, $tt\bar{t}$ and $\bar{t}t\bar{t}$. We are discarding the $t\bar{t}j$ final state because of the 2LSS and 3L selection criteria. On the other hand, $t\bar{t}Z'$ can yield $t\bar{t}tj$, $\bar{t}t\bar{t}j$ and $t\bar{t}t\bar{t}$ final states.

If we want events enriched with leptonic charge asymmetry and b-jets, we need $tZ'$ to decay mostly to three top quarks while being as charge asymmetric as possible. That is, we need $BR(Z' \to t\bar{t}) > BR(Z' \to tj + \bar{t}j)$. As we consider relatively low $M_{Z'}$

masses, we need $g_{tt} \gg g_{ut}$ to avoid phase-space suppression. If $M_{Z'} < 2m_t$, we consider the three-body decay $Z' \to tW^-\bar{b}$, $\bar{t}W^+b$ we detail in Subsection 1.2.2. From Fig. 19 we see that for the benchmarks points we display, when $M_{Z'}$ is large enough the $t\bar{t}$ decay mode dominates. When combined with $tZ'$ and $t\bar{t}Z'$ production, these decays produce three- and four-top-quark final states.

After this overview of the basic phenomenology of the proposed model, we turn to studying its effect on the relevant observables, and how these observables determine the region in its parameter space most compatible with the experimental results.

## 4.3 EXPERIMENTAL IMPRINTS AND MODEL TUNING

In this Section we study how the $Z'$ model detailed in Section 4.2 is probed by different existing experimental results. A $Z'$ as in Eq. 58 is constrained by both high- and low-energy observables. In light of recent experimental results from $t\bar{t}H$ and $t\bar{t}t\bar{t}$ searches in ATLAS and CMS, we are specially interested in same sign dilepton and multilepton plus b-jets final states. In particular, final states with non-zero total leptonic charge are highly sensitive to $Z'$. Taking this into account, we detail in Subsection 4.3.1 how on-shell $Z'$ production in association either with a single top quark ($tZ'$ or $\bar{t}Z'$) or with a top quark pair ($t\bar{t}Z'$) could explain the need for $t\bar{t}W^\pm$ re-scaling to account for tensions in data in recent $t\bar{t}H$ results [151], while yielding interesting signatures in 4-top searches. In Subsection 4.3.2 we test whether the parameter space indicated by the same sign dilepton and multilepton and high b-jet multiplicity searches is safe to other observables that could be affected by our model. The more relevant observables we consider are $D^0 - \bar{D}^0$-meson mixing, top-quark rare decays through flavour-changing neutral currents (FCNC), top-quark pair production, same sign top-quark pair production, resonant tj production in $t\bar{t}$+jets events, and $Z'$ radiative production $tZ'j$. Other processes involving resonant $Z'$ production, such as single $Z'$, $Z'j$ and single top quark production $tj + \bar{t}j$, are absent due to having neglected the $g_{uu}$ and $g_{cc}$ couplings.

### 4.3.1 *Fits to experimental data*

We study how our phenomenological $Z'$ model detailed in Section 4.2 accommodates recent results reported for $t\bar{t}H$ production while avoiding constraints and yielding potentially interesting results in $t\bar{t}t\bar{t}$ searches. Regarding $t\bar{t}H$, we focus on the ATLAS preliminary results reported in Ref. [151], although results from CMS [152] are consistent with our conclusions. Regarding 4-top searches we consider the results reported by ATLAS in Ref. [25], having corroborated that the results obtained are compatible with the combination of the ATLAS and CMS [26] results. Both searches target same sign dilepton and multilepton processes but their channel definitions are not the same and the reconstructed objects, both leptons and jets, have different kinematic cuts and tagging efficiencies, which we take into account in our study.

For the case of $t\bar{t}H$, the results reported in Ref. [151] are particularly interesting because of the difficulties reported when dealing with the irreducible $t\bar{t}W^\pm$ background. Figure 2 of Ref. [151] highlights the need for a missing charge asymmetric contribution to match the data, which yields a normalization factor for $t\bar{t}W^\pm$ larger than one. This is consistent with other measurements, for example Refs. [149, 150], and has motivated a push for higher theoretical accuracy in the $t\bar{t}W^\pm$ calculations [24, 173–179], which

nevertheless have not fully explained the discrepancy between the expected and the observed $t\bar{t}W^{\pm}$ event yields.

If we treat the $t\bar{t}W^{\pm}$ background as well modelled, and thus constrained to have a normalization factor consistent with unity within the uncertainty in its theoretical cross-section, we are faced with charge asymmetric anomalous events with high b-jet multiplicity. Our $Z'$ model is designed to accommodate these two features. A recent example of BSM effects in $t\bar{t}W$ can be found in Ref. [155], where, in contrast to a resonant BSM physics model, they study an Effective Field Theory in the top-quark sector.

As detailed in Subsection 4.2.1, we study $Z'$ production in association either with a single top quark or with a top-quark pair. These processes, $tZ'$ and $t\bar{t}Z'$, along with a considerable $Z' \to t\bar{t}$ Branching Ratio achieved with a suitable coupling hierarchy, provide the necessary same sign dilepton and multilepton signatures, with both charge asymmetry and high b-jet multiplicity.

As we consider three-top-quarks production and four-top-quarks production, we need to make sure that our results are compatible with 4-top limits. Even if four-top-quarks production will be mostly sensitive to $(g_{tt})^4$, three-top-quarks production is proportional to $(g_{ut})^2(g_{tt})^2$ and will introduce a charge asymmetry.

To see how well our model can agree with the data, we obtain the event yields expected from $tZ'$ and $t\bar{t}Z'$ in the different reported bins in Fig. 2 of Ref. [151], and the expected events in 4-top searches with the selection criteria of Ref. [25] as a function of $(g_{ut}, g_{ct}, g_{tt})$ for different values of $M_{Z'}$. To do this, we have implemented the $Z'$ model in Eq. 58 using `Feynrules` [14] and then simulated a fixed set of points through `MadGraph5_aMC@NLO` [12] for production and decay of $tZ'$ and $t\bar{t}Z'$. The signal events have been generated using a leading-order matrix element and the NN23LO1 PDF set [121], and have been processed through `Pythia 8` [16] for the modelling of parton showering and hadronization, as well as through a simulation of the detector response as implemented in `Delphes 3` [18]. We employ tuned parameters for `Pythia 8` and `Delphes 3`. For the former, we use the Monash tune [180] with a few changes aimed to reproduce the $t\bar{t}W^{\pm}$ $N_j$ distribution as faithfully as possible. We detail the changed parameters in Table 6.

In the `Delphes 3` case, we tune a `Delphes 3` card for the $t\bar{t}H$ search [151] in order to match the expected $t\bar{t}W^{\pm}$ event yields with the reference cross-section $\sigma = 727$ fb. We use almost the same `Delphes 3` card for the 4-top search [25], with the only difference being the b-tagging efficiency. The detailed features of the first `Delphes 3` card are:

- We set the electron identification efficiency as

$$\begin{aligned}
0.0 \quad &\text{for} \quad p_T \leqslant 10.0 \\
0.80 \cdot (0.01 \cdot (p_T - 20.0) + 0.65) \quad &\text{for} \quad 10.0 < p_T \leqslant 45.0 \text{ and } |\eta| \leqslant 1.5 \\
0.75 \cdot (0.01 \cdot (p_T - 20.0) + 0.65) \quad &\text{for} \quad 10.0 < p_T \leqslant 45.0 \text{ and } 1.5 < |\eta| \leqslant 2.5 \\
0.80 \cdot 0.90 \quad &\text{for} \quad 45.0 < p_T \text{ and } |\eta| \leqslant 1.5 \\
0.75 \cdot 0.90 \quad &\text{for} \quad 45.0 < p_T \text{ and } 1.5 < |\eta| \leqslant 2.5 \\
0.0 \quad &\text{for} \quad |\eta| > 2.5
\end{aligned}$$

| Parameter | Default | Value |
|---|---|---|
| TimeShower:alphaSvalue | 0.1365 | 0.127 |
| SpaceShower:pToRef | 2.0 | 1.56 |
| SigmaProcess:alphaSvalue | 0.130 | 0.140 |
| SpaceShower:pTmaxFudge | 1.0 | 0.91 |
| SpaceShower:pTdampFudge | 1.0 | 1.05 |
| SpaceShower:alphaSvalue | 0.1365 | 0.127 |
| BeamRemnants:primordialKThard | 1.8 | 1.88 |
| MultipartonInteractions:pToRef | 2.28 | 2.09 |
| MultipartonInteractions:alphaSvalue | 0.130 | 0.126 |
| ColourReconnection:range | 1.8 | 1.71 |

Table 6: This table lists the `Pythia 8` parameters changed from their Monash Tune default values [180].

- We set the muon identification efficiency as

$$\begin{aligned}
0.0 \quad &\text{for} \quad p_T \leqslant 10.0 \\
0.98 \cdot (0.006 \cdot (p_T - 20.0) + 0.80) \quad &\text{for} \quad 10.0 < p_T \leqslant 45.0 \text{ and } |\eta| \leqslant 1.5 \\
0.99 \cdot (0.006 \cdot (p_T - 20.0) + 0.80) \quad &\text{for} \quad 10.0 < p_T \leqslant 45.0 \text{ and } 1.5 < |\eta| \leqslant 2.5 \\
0.98 \cdot 0.95 \quad &\text{for} \quad 45.0 < p_T \text{ and } |\eta| \leqslant 1.5 \\
0.99 \cdot 0.95 \quad &\text{for} \quad 45.0 < p_T \text{ and } 1.5 < |\eta| \leqslant 2.5 \\
0.0 \quad &\text{for} \quad |\eta| > 2.5
\end{aligned}$$

- We set the electron and muon isolation parameters as

$$\begin{aligned}
R &= 0.3 \\
p_T^{min} &= 1.0 \\
I_{min} &= 0.2
\end{aligned}$$

- We cluster jets with the anti-kt jet clustering algorithm [181], with $\Delta R = 0.4$ and $p_T^{min} = 25.0$
- We set the jet energy scale as

$$\sqrt{\frac{(3.0 - 0.2 \cdot |\eta|)^2}{p_T} + 1.0}$$

- We set the b-tagging efficiencies as

$$\begin{aligned}
\text{Default misidentification rate} \quad &: \quad \frac{440.0}{313.0} \cdot (0.002 + 7.3 \cdot 10^{-06} \cdot p_T) \\
\text{c-jet misidentification rate} \quad &: \quad 0.20 \cdot \tanh(0.02 \cdot p_T) \cdot \frac{1}{1 + 0.0034 \cdot p_T} \\
\text{b-jet efficiency} \quad &: \quad 0.80 \cdot \tanh(0.003 \cdot p_T) \cdot \frac{30}{1 + 0.086 \cdot p_T}
\end{aligned}$$

- We modify the τ-tagging module to match the medium working point in Ref. [151]

```
module TrackCountingTauTagging TauTagging {

  set ParticleInputArray Delphes/allParticles
  set PartonInputArray Delphes/partons
  set TrackInputArray TrackMerger/tracks
  set JetInputArray JetEnergyScale/jets

  set DeltaR 0.2
  set DeltaRTrack 0.2

  set TrackPTMin 1.0

  set TauPTMin 20.0
  set TauEtaMax 2.5

  # instructions: {n-prongs} {eff}

  # 1 - one prong efficiency
  # 2 - two or more efficiency
  # -1 - one prong mistag rate
  # -2 - two or more mistag rate

  set BitNumber 0

  add EfficiencyFormula {1} {0.55}
  add EfficiencyFormula {2} {0.40}
  add EfficiencyFormula {-1} {0.02}
  add EfficiencyFormula {-2} {0.002}

}
```

For the 4-top `Delphes 3` card we modify the b-tagging efficiencies to

$$\begin{aligned}
\text{Default misidentification rate} &: \frac{440.0}{313.0} \cdot (0.002 + 7.3 \cdot 10^{-06} \cdot p_T) \\
\text{c-jet misidentification rate} &: \frac{8.1}{4.0} \cdot 0.20 \cdot \tanh(0.02 \cdot p_T) \cdot \frac{1}{1 + 0.0034 \cdot p_T} \\
\text{b-jet efficiency} &: \frac{77.0}{70.0} \cdot 0.80 \cdot \tanh(0.003 \cdot p_T) \cdot \frac{30}{1 + 0.086 \cdot p_T}
\end{aligned}$$

This `Delphes 3` card allows us to recover the expected yields for $t\bar{t}W^{\pm}$, 4-top production, $t\bar{t}H$ and $t\bar{t}Z$ reported in Table 3 of Ref. [25] to a very good degree of approximation.

After simulating the events, we implement the event selection cuts and obtain the event yields. Each search has a different event selection for each channel as detailed in Tables 4 and 5. Additionally, we require at least 4 jets in the 2LSS $t\bar{t}H$ selection and at least 3 b-jets in the 2LSS and 3L 4-top selection. The former modification is needed

to compare our results to Fig. 2 in Ref. [151] and the latter is needed to obtain a signal-enhanced selection similar to the one defined by the use of the BDT in Ref. [25]. We also incorporate specific trigger selection efficiencies for each leptonic channel to the `ROOT` [19] code used to analyse the `Delphes` output. The simulated signal samples have been normalized using k-factors obtained from simulating similar events to NLO with the same set-up, and which are consistent with those in the literature [182].

This method of generating events is computationally expensive and thus makes it challenging to generate a large set of events to constrain the parameter space of our model. However, we are able to generate a large set of points in parameter space from a relatively small set of Monte Carlo generated points by observing that the expected number of events from a given process in an analysis channel, $N^{\text{channel}}_{\text{process}}$, can be parametrized as follows[1]:

$$\begin{aligned}
N^{ch}_{tZ'} &= A^{ch}_1 \cdot BR(Z' \to t\bar{u} + \bar{t}u) \cdot (g_{ut})^2 + A^{ch}_2 \cdot BR(Z' \to t\bar{c} + \bar{t}c) \cdot (g_{ut})^2 \\
&\quad + A^{ch}_3 \cdot BR(Z' \to t\bar{t}) \cdot (g_{ut})^2 + A^{ch}_4 \cdot BR(Z' \to t\bar{u} + \bar{t}u) \cdot (g_{ct})^2 \\
&\quad + A^{ch}_5 \cdot BR(Z' \to t\bar{c} + \bar{t}c) \cdot (g_{ct})^2 + A^{ch}_6 \cdot BR(Z' \to t\bar{t}) \cdot (g_{ct})^2, \\
N^{ch}_{t\bar{t}Z'} &= B^{ch}_1 \cdot BR(Z' \to t\bar{u} + \bar{t}u) \cdot (g_{tt})^2 + B^{ch}_2 \cdot BR(Z' \to t\bar{c} + \bar{t}c) \cdot (g_{tt})^2 \\
&\quad + B^{ch}_3 \cdot BR(Z' \to t\bar{t}) \cdot (g_{tt})^2,
\end{aligned} \tag{59}$$

where all the coefficients $A^{ch}_i, B^{ch}_i$ are functions of $M_{Z'}$ and absorb the acceptance of the channel and the cross-section for the specific process normalized to the corresponding coupling set to unity. The channels we consider are the different bins of Fig. 2 of Ref. [151] and the 4-top event yield with $Q > 0$ and with $Q < 0$ with the selection criteria of Ref. [25]. We obtain the $A^{ch}_i$ and $B^{ch}_i$ with Weighted Least Squares, where the uncertainty of each measurement is due to the Monte Carlo finite sampling, and with them we generate arbitrary points in the parameter space.

The likelihood fit to the eight bins of Fig. 2 of Ref. [151] is performed using the `HistFitter` package [124], which relies on `RooFit` [125] and the minimization algorithms from `MINUIT` [126]. Additionally, systematic uncertainties affecting the overall normalization of the SM processes are included in the fit: 20% uncertainty is assigned to $t\bar{t}H$, $t\bar{t}W$, $t\bar{t}Z$, and signal, and 50% uncertainty to diboson, while the various fake lepton components have uncertainties assigned corresponding to the normalisation factor precision reported in the ATLAS result.

In Figs. 20-22 we compute the impact of different points in parameter space for different $M_{Z'}$ masses on the two experimental analysis. In the left column of these plots, we obtain the point that minimizes the Negative Log-Likelihood (NLL) [5] for the reported observed events in Fig. 2 of the $t\bar{t}H$ search [151], considering the pre-fit SM background contributions. After finding this minimum, we plot the 1 and 2 standard deviations (s.d.) regions and the goodness-of-fit contour lines [183]. We include in this same left column the contour lines indicating the 4-top BSM and SM to SM only event ratio, $N^{\text{4-top}}_{SM+BSM}/N^{\text{4-top}}_{SM}$. These events need to pass the 4-top-like selection cuts described in Table 5 and include SM four-top-quarks and BSM $t\bar{t}t + t\bar{t}\bar{t}$, $t\bar{t}tj + t\bar{t}\bar{t}j$ and four-top-quarks processes. We can see the interplay between a good $t\bar{t}H$ fit, which requires asymmetry (non-negligible $g_{ut}$) but also a high b-jet multiplicity

1 This parametrization neglects non-resonant $Z'$ effects.

(large $g_{tt}$), and the 4-top fit (not so large $g_{tt}$). We combine both measurements in the right column where we minimize the $t\bar{t}H$+ 4-top data

$$\mathrm{NLL}(t\bar{t}H) + \frac{(N^{4\text{-top}}_{SM+BSM}/N^{4\text{-top}}_{SM} - N^{4\text{-top}}_{obs}/N^{4\text{-top}}_{SM})^2}{\sigma^2}$$

For 4-top we consider two $N^{4\text{-top}}_{obs}/N^{4\text{-top}}_{SM}$ possibilities: the reported ATLAS [25] value alone $N^{4\text{-top}}_{obs}/N^{\text{four-tops}}_{SM} = 2.0^{+0.8}_{-0.6}$ and in combination with the CMS [26] reported value $N^{\text{four-tops}}_{obs}/N^{\text{four-tops}}_{SM} = 1.1 \pm 0.5$, which yields an average value of $N^{\text{four-tops}}_{obs}/N^{\text{four-tops}}_{SM} = 1.4 \pm 0.3$. In all cases we find that the best fit points are compatible and we chose to focus on the ATLAS result to tune our simulation and obtain further information. We plot the 1 s.d. and 2 s.d. regions for the ATLAS value and we showcase how the $Z'$ introduces an imbalance in the ratio of yields of three- and four-top-quarks events with positive and negative total leptonic charge

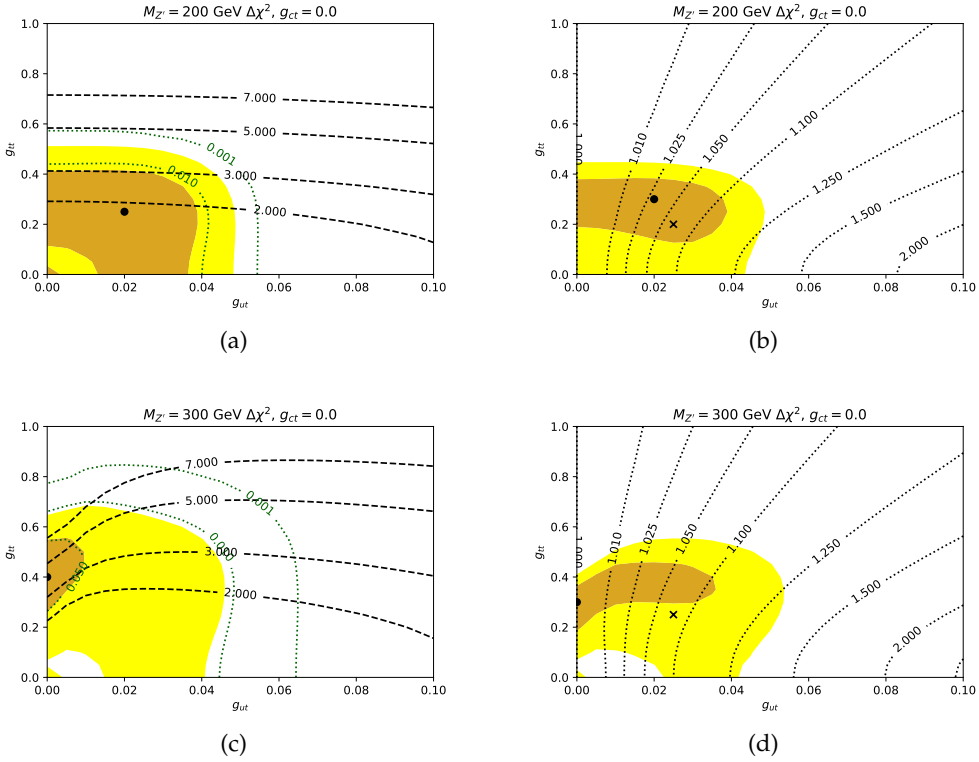
$$r(4t) = \frac{N^{4\text{-top},Q>0}_{SM+BSM}}{N^{4\text{-top},Q<0}_{SM+BSM}} \frac{N^{4\text{-top},Q<0}_{SM}}{N^{4\text{-top},Q>0}_{SM}}, \tag{60}$$

in the 4-top search. Observe that the ATLAS 4-top analysis [25] does not distinguish three- and four-top-quarks final states and therefore this imbalance is also induced by diagrams of the kind of Fig. 18a because of the $u$ quark in the initial state and, as expected, grows with $g_{ut}$. We plot the contour levels for $r(4t)$ in the figures since it provides a qualitative insight on the behaviour of the charge imbalance coming from the $g_{ut}$-mediated BSM contributions in a 4-top-like selection. We also observe that this parameter is not currently reported by the experimental collaborations, and could provide clues of BSM contributions.

To better identify interesting features, we plot in Fig 20 the region $M_{Z'} < 2m_t$ and $g_{ct} = 0$; in Fig. 21 we study $M_{Z'} > 2m_t$ and $g_{ct} = 0$; and in Fig. 22 we explore the $g_{ct} \neq 0$ region while fixing $g_{tt} = 0.2$ and 0.4. In the following paragraphs we discuss each one of these figures in detail. Although we find many points in parameter space compatible with the experimental results, we observe that $M_{Z'} = 400$ GeV has a slightly better accordance with the data.

For the $M_{Z'} < 2m_t$ case shown in Fig. 20, the 1 s.d. regions in all plots include $g_{ut} \approx 0$. This corresponds to no charge asymmetry and is indicative of the fact that none of the points in the parameter space are a particularly good fit for the data in this region, as quantified by the poor goodness-of-fit not being significantly better than that of the SM hypothesis ($g_{ut} = g_{ct} = g_{tt} = 0$). This is because $Z' \to t\bar{t}$ is suppressed and thus the events providing charge asymmetry with large b-jet multiplicity are suppressed as well. This leaves $Z' \to t\bar{u} + \bar{t}u$ as the dominant decay mode, worsening the fit. When comparing the right column ($t\bar{t}H$ and 4-top) to the left column (only $t\bar{t}H$), one should keep in mind that yields in 4-top searches are obtained mainly through the three-top-quarks production diagrams, which are proportional to $(g_{tt})^2 \mathrm{BR}(Z' \to t\bar{u} + \bar{t}u)$ and $(g_{ut})^2 \mathrm{BR}(Z' \to t\bar{t}^* + t^*\bar{t})$. For $M_{Z'} = 200$ GeV, the second process is negligible and therefore the effect of incorporating 4-top to the NLL is to constrain $g_{tt}$. Whereas for $M_{Z'} = 300$ GeV, there is a slight opening of $Z' \to t\bar{t}^* + t^*\bar{t}$ and thus the 1 s.d. allowed region enlarges to the medium $g_{ut}$ region from Fig. 20c to 20d.

The goodness-of-fit to $t\bar{t}H$ data increases when we consider $M_{Z'} > 2m_t$, as shown in Fig. 21. In fact, for these $M_{Z'}$ values the $Z' \to t\bar{t}$ channel opens up and diagrams of the kind shown in Fig. 18a provide same sign and multilepton events with charge asymmetry and large b-jet multiplicity, which are crucial to improve the fit. This




(a) (b) (c) (d)

Figure 20: Fits to data for $M_{Z'} < 2m_t$ in the $(g_{ut}, g_{tt})$ plane. Left column: ellipses correspond to 1 s.d. and 2 s.d. from the $t\bar{t}H$ NLL minimum with the green dotted curve being the goodness-of-fit and the black dashed curve corresponding to $N^{4\text{-top}}_{SM+BSM}/N^{4\text{-top}}_{SM}$. Right column: ellipses correspond to 1 s.d. and 2 s.d. from the $t\bar{t}H+$ 4-top NLL minimum with the black dotted curve corresponding to the charge imbalance ratio of yields $r(4t)$ defined in Eq. 60. The $t\bar{t}H+$ ATLAS and CMS average 4-top NLL minimum is shown with a black cross.

important improvement can be explicitly seen by comparing the goodness-of-fit (green dotted contours) between the $M_{Z'} < 2m_t$ (Fig. 20) and $M_{Z'} > 2m_t$ (Fig. 21) results in the left plots. We can see in the left plots that as $M_{Z'}$ increases from 400 GeV to 600 GeV, the $tZ'$ and $t\bar{t}Z'$ production cross-sections decrease and therefore larger couplings are needed for the best-fit regions. We see however in these plots that the best-fit regions are in potential tension with 4-top production cross-section (black dashed contours), which indicate a preference for lower values for $g_{tt}$. In the right plots of Fig. 21 we include 4-top data to the NLL and we observe that the best-fit point has a noticeable lower $g_{tt}$ and similar $g_{ut}$. This is because 4-top searches are more sensitive to $g_{tt}$ than to $g_{ut}$. We do not consider masses above 600 GeV because we find that larger masses would yield $H_T$ distributions that enter into conflict with those reported in 4-top analyses, as discussed below in Fig. 27.

We plot the $g_{ct} \neq 0$ cases in Fig. 22 for $M_{Z'} = 400$ GeV and two values of $g_{tt}$. In all cases we see that the best-fit point is in $g_{ct} = 0$, which indicates that the main handles to accommodate the data are $g_{ut}$ and $g_{tt}$. Nevertheless, we observe that $g_{ct} \neq 0$ is allowed at the 1 s.d. level in a large part of parameter space. We show in blue the region disfavoured by $D^0 \leftrightarrow \bar{D}^0$ mixing as detailed in Subsection 4.3.2, which limits the $g_{ut}$ and $g_{ct}$ couplings.

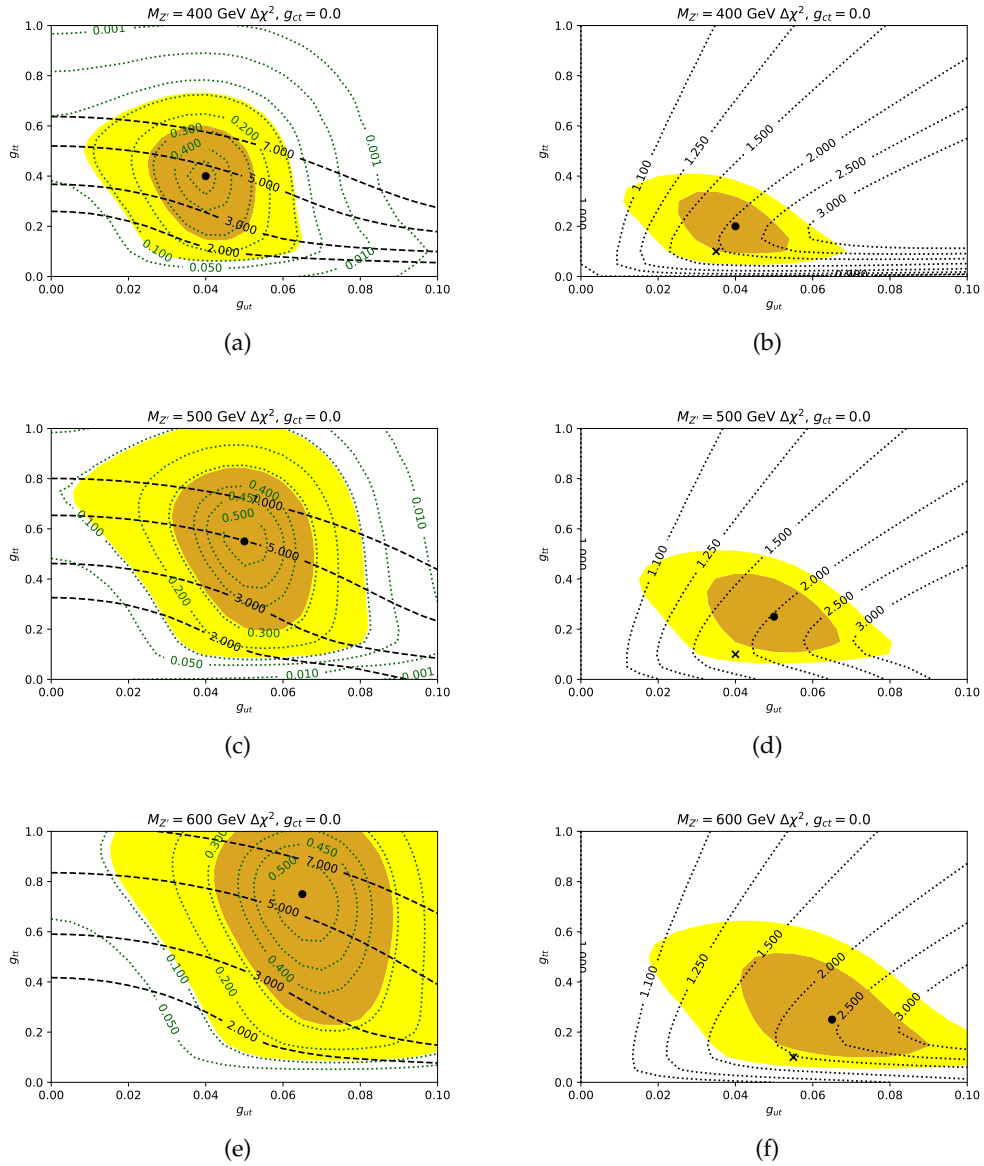
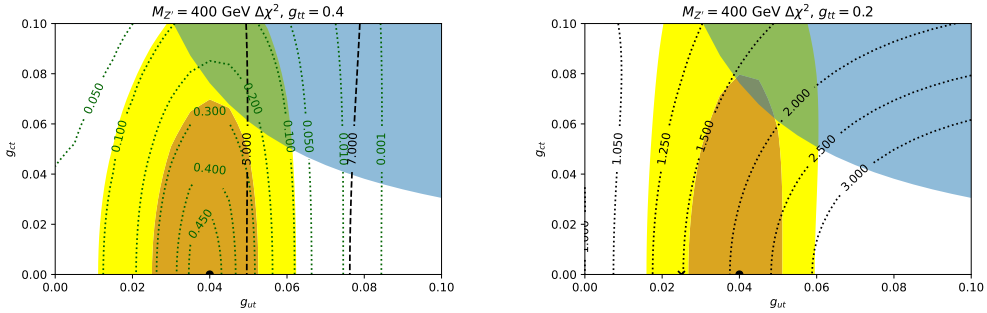



Figure 21: Idem as in Fig. 20, but for $M_{Z'} > 2m_t$. In this case, the opening of the decay $Z' \to t\bar{t}$ enhances the channel $ug \to tZ' \to t\bar{t}t$ which induces a charge imbalance that favours considerably the fit in $t\bar{t}H$. This can be appreciated in the goodness-of-fit (green dotted), which is considerably better in this case than for those in Fig. 20.

From these fits to the data we can conclude that the $Z'$ is not only compatible with the $t\bar{t}H$ and the 4-top data, but also in many regions is more compatible that the SM. The $tZ'$ process can provide the necessary charge asymmetry and b-jet multiplicity while still being hidden in 4-top production. To summarize how selecting a good benchmark point in parameter space reduces the tension in the reported results in $t\bar{t}H$ (Fig. 2 in Ref. [151]), we show in Fig. 23 how these results are modified if the $Z'$ BSM is added to the SM yields. We consider the best-fit point for $t\bar{t}H$ data only for $M_{Z'} = 400$ GeV: $g_{ut} = 0.04$, $g_{ct} = 0.0$ and $g_{tt} = 0.4$. For comparison, we show the equivalent post-fit plot in the case where only SM processes are considered in the fit. In both fits, the same systematic model is used for the SM processes.




(a) (b)

Figure 22: Idem as in Fig. 21, but for different points in the $(g_{ut}, g_{ct})$ plane. The blue regions correspond to regions in parameter space disfavoured by $D^0 \leftrightarrow \bar{D}^0$ mixing as detailed in Section 4.3.2.

Although all masses above $2m_t$ provide good fits, we consider the best-fit point for $t\bar{t}H$ and 4-top data corresponding to $M_{Z'} = 400$ GeV and couplings $g_{ut} = 0.04$, $g_{ct} = 0.0$ and $g_{tt} = 0.2$ as our benchmark point for further studies. As we show in Section 4.4, this mass has the interesting feature of having an $H_T$ distribution similar to, although slightly softer than, that of SM 4-top production.

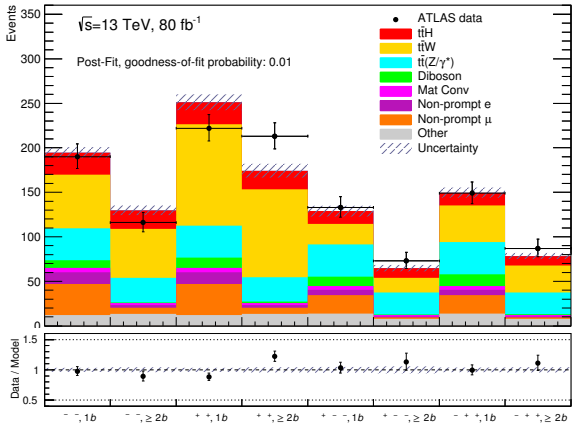
At last, it is interesting to observe that the preferred region in parameter space by the fit indicates that Z′ associated production either with a single top-quark or with a top-quark pair can mimic $t\bar{t}W^{\pm}$ in the $t\bar{t}H$ search and 4-top in the 4-top search. In Section 4.4, we study kinematic distributions and properties of the new signals compared to the main SM background processes in order to find ways to break this degeneracy.

### 4.3.2 *Constraints from other observables*

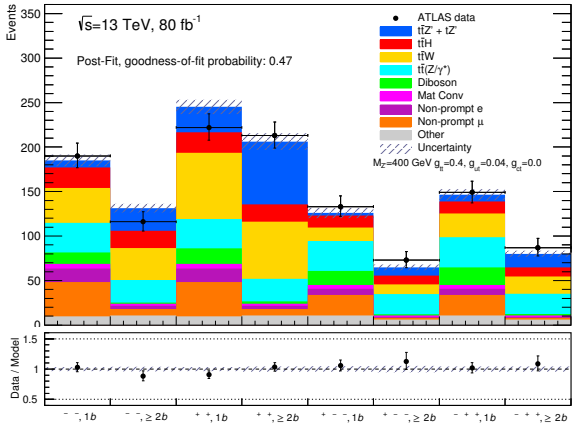
In this Subsection we analyse how other observables besides $t\bar{t}H$ and 4-top constrain the previously studied parameter space, namely $0 \leqslant g_{ut} \leqslant 0.1$, $0 \leqslant g_{ct} \leqslant 0.1$ and $0 \leqslant g_{tt} \leqslant 1.0$; and $200 \leqslant M_{Z'} \leqslant 600$ GeV.

As a low-energy physics phenomena, D-meson mixing is particularly sensitive to any BSM effects. In particular, since virtual Z′ and top quarks can contribute to $D^0 \leftrightarrow \bar{D}^0$ mixing at loop level, it is a sensitive observable to our model. Since we consider $g_{uc} = 0$, there is no tree level contribution to D-meson mixing. The Lorentz structure in Eq. 58 provides a contribution to D-meson mixing that is translated to the D-meson mass difference. We proceed to compute this one-loop contribution to $D^0 - \overline{D^0}$. We observe that the general tree level computation has already been performed in Ref. [184], whereas the one-loop contribution with equal couplings for both chiralities $g^L_{ut,ct} = g^R_{ut,ct}$ can be found in Ref. [185]. We thus follow and expand the $|\Delta C| = 2$ calculations in previous references to the one-loop case with different couplings for each chirality $g^{L,R}_{ut,ct}$.

The relevant Feynman diagrams to calculate the one-loop contribution are shown in Fig. 24. The corresponding amplitude is




(a)




(b)

Figure 23: Comparison between ATLAS data [151] and a) SM background prediction b) BSM signal ($tZ' + t\bar{t}Z'$) plus SM background prediction for the event yields in 2LSS and 3L channels in categories based on the total lepton charge and the b-jet multiplicity. The BSM signal corresponds to a $M_{Z'} = 400$ GeV and couplings $g_{ut} = 0.04$, $g_{ct} = 0.0$ and $g_{tt} = 0.4$, which minimizes the NLL for the ATLAS data [151] for $M_{Z'} = 400$ GeV. In both cases, we show the SM background prediction with fitted NP to the data although the fitted values differ due to the presence of BSM events in b).

$$\mathcal{M} = 2\int \frac{d^4k}{(2\pi)^4} \frac{\left[\overline{\mathcal{P}_{ut}u}\gamma^\lambda(\not{k}+m_t)\gamma^\nu\mathcal{P}_{ct}c\right]\left[\overline{\mathcal{P}_{ut}u}\gamma^\nu(\not{k}+m_t)\gamma^\lambda\mathcal{P}_{ct}c\right]}{(k^2-m_t^2)^2\,(k^2-M_{Z'}^2)^2} \tag{61}$$

where

$$\mathcal{P}_{ut} = g_{ut}^L P_L + g_{ut}^R P_R \tag{62}$$

$$\mathcal{P}_{ct} = g_{ct}^L P_L + g_{ct}^R P_R \tag{63}$$

and $P_{L,R} = (1 \mp \gamma^5)/2$. We work in the low-energy limit where the u and c quark fields have no $k^\mu$ dependence.

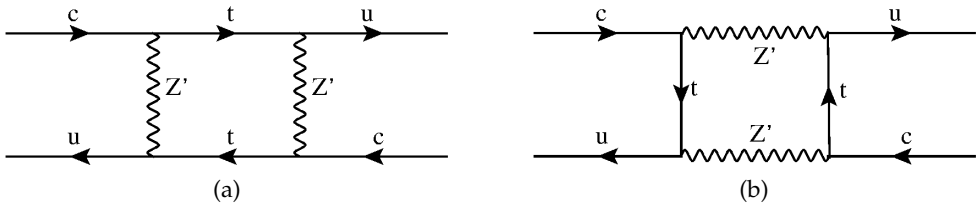



Figure 24: One-loop box Feynman diagrams contributing to $D^0 - \overline{D^0}$ mixing due to a generic $Z'$ with non-zero off-diagonal couplings to the top quark.

In order to perform the integral on the loop momentum $k$ in Eq. 61 we observe that by symmetry arguments any integral with odd number of $k$'s in the numerator vanishes. We are then left with the two relevant terms proportional to $\not{k}\,\not{k}$ and $m_t^2$. With the help of dimensional regularization going to $d$ dimensions, we can use the identity

$$\frac{1}{A^m B^n} = \frac{\Gamma(m+n)}{\Gamma(m)\Gamma(n)} \int_0^\infty d\lambda \frac{\lambda^{m-1}}{(\lambda A + B)^{m+n}} \tag{64}$$

to obtain the solvable d-dimensional integrals [1]

$$\begin{aligned} \int \frac{d^d k}{(2\pi)^d} \frac{1}{(k^2-\Delta)^n} &= \frac{(-1)^n i}{(4\pi)^{d/2}} \frac{\Gamma(n-\frac{d}{2})}{\Gamma(n)} \left(\frac{1}{\Delta}\right)^{n-\frac{d}{2}}, \\ \int \frac{d^d k}{(2\pi)^d} \frac{k^\mu k^\nu}{(k^2-\Delta)^n} &= \frac{(-1)^{n-1} i}{(4\pi)^{d/2}} \frac{g^{\mu\nu}}{2} \frac{\Gamma(n-\frac{d}{2}-1)}{\Gamma(n)} \left(\frac{1}{\Delta}\right)^{n-\frac{d}{2}-1}, \end{aligned} \tag{65}$$

and setting $d = 4$ in the end. By performing the d-dimensional integration and then the $\lambda$ integration, we obtain an expression for $\mathcal{M}$ without any remaining integrals. We can then use Fierz transformations to write $\mathcal{M}$ in terms of the four-fermion effective operators $Q_{1\ldots8}$ defined in Ref. [184] and write the effective Lagrangian $\mathcal{L}_{eff}$ induced by the FCNC $Z'$ in terms of these operators. This allows us to use the *modified vacuum saturation* hypothesis as defined in Ref. [184] to obtain $\langle Q_i \rangle = \langle \overline{D^0} | Q_i | D^0 \rangle$ and with them an explicit expression for $\langle \overline{D^0} | \mathcal{L}_{eff} | D^0 \rangle$. Utilizing the definition for

$$\Delta M_D = -\frac{1}{M_D} |\langle \overline{D^0} | \mathcal{L}_{eff} | D^0 \rangle| \tag{66}$$

we obtain the one-loop D-meson mixing parameter for chiral off-diagonal couplings with the top quark

$$\begin{aligned} \Delta M_D = \frac{f_D^2 M_D B_D x}{64\pi^2 M_{Z'}^2} \Bigg[ & f(x) \left( -\frac{8}{3} g_{ut}^{L\,2} g_{ct}^{L\,2} - \frac{8}{3} g_{ut}^{R\,2} g_{ct}^{R\,2} + \frac{80}{3} g_{ut}^L g_{ct}^L g_{ut}^R g_{ct}^R \right) \\ & + g(x) \left( \frac{2}{3} g_{ut}^{L\,2} g_{ct}^{R\,2} + \frac{2}{3} g_{ut}^{R\,2} g_{ct}^{L\,2} - \frac{14}{3} g_{ut}^L g_{ct}^L g_{ut}^R g_{ct}^R \right) \Bigg]. \end{aligned} \tag{67}$$

where $x = (M_{Z'}/m_t)^2$, $f_D = 222.6$ MeV [186] is the $D^0$-meson decay constant, $B_D \approx 1$ [185, 187] is the bag model parameter, and

$$f(x) = \frac{1}{2} \frac{1}{(1-x)^3} [1 - x^2 + 2x \log x] \tag{68}$$

$$g(x) = \frac{2}{(1-x)^3} [2(1-x) + (1+x) \log x] \tag{69}$$

are the functions that appear when integrating in $\lambda$ in the above scheme. As it can be easily seen, the results in Ref. [185] are recovered if we set the left- and right-handed couplings to be equal. We obtain the one-loop contribution to D-meson mixing of our $Z'$ model by setting $g_{ut}^L = g_{ct}^L = 0$, $g_{ut}^R = g_{ut}$ and $g_{ct}^R = g_{ct}$, obtaining

$$\Delta M_D^{BSM} = -\frac{f_D^2 M_D B_D x}{64\pi^2 M_{Z'}^2}\left(\frac{8}{3} f(x)\,(g_{ut})^2\,(g_{ct})^2\right). \tag{70}$$

Although in principle one should compare the experimental value to the predicted value due to SM and BSM contributions, a detailed knowledge of $\Delta M_D^{SM}$ is currently lacking [5, 188, 189]. Therefore, as a naive estimation, we require that the BSM contribution to $\Delta M_D$ is smaller than the uncertainty in its measurement:

$$\Delta M_D^{BSM} \lesssim \delta\left(\Delta M_D^{exp}\right). \tag{71}$$

Using current available data [5] we obtain that the above constraints are translated into the model parameters as

$$g_{ut}\, g_{ct} < 2.0 \times 10^{-3} \text{ to } 4.5 \times 10^{-3} \tag{72}$$

for $M_{Z'}$ ranging from 200 GeV to 600 GeV, respectively. These bounds are plotted as blue regions in Fig. 22.

High-energy collider physics is sensitive both to inclusive on-shell production of the $Z' + X$ and to non-resonant behaviour, such as the re-scaling of different top-physics cross-sections [28, 64, 166, 169]. As mentioned before, taking $g_{uu}, g_{cc} = 0$ makes our model insensitive to many of them. Other effects are greatly diminished.

Measuring top-quark rare decays opens a window to BSM effects. $Z'$-induced top-quark rare decays have been studied in Ref. [66] and can be confronted to the experimental bounds set in Refs. [67–70]. Using the same matrix elements we can produce the relevant decay Branching Ratios for the model described in Eq. 58. After adding the smaller SM contributions, we find that rare decays Branching Ratios are below current limits, which are of $\mathcal{O}(10^{-5})$. The main difference between our model and those considered in Ref. [66] is the absence of $g_{uu}$ and $g_{cc}$ couplings and of left-handed couplings, which effectively reduces the $Z'$-induced rare decays and allows us to go to smaller masses.

As mentioned above, in high-energy collider searches our model yields no $pp \to Z'$ production. In $t\bar{t}$ production, it can only contribute through the t-channel and cannot be probed by $t\bar{t}$ resonance searches. The only constraints coming from $t\bar{t}$ are then those on the inclusive cross-section, where we find that $Z'$ effects are well below the current experimental bounds [190, 191] for the parameter space we are exploring, being at most 45 fb for $g_{ut} = 0.1$, $g_{ct} = 0.1$, $g_{tt} = 1.0$ and $M_{Z'} = 200$ GeV.

Our $Z'$ model is also sensitive to same sign top-quark pair searches, both to prompt same sign top-quark pair production mediated by a $Z'$ in the t-channel and to same sign top-quark production with an associated jet. The former is sensitive to $(g_{ut})^4$, $(g_{ct})^4$, and $(g_{ut})^2(g_{ct})^2$, the dominant process being $uu \to tt$ due to the PDF imbalance between the up quark and the others. The latter corresponds to the same $tZ'$ production mechanism we discuss in Subsection 4.2.1, but we now select events where the $Z'$ decays to $t\bar{u} + t\bar{c}$. Current experimental limits from prompt same sign top-quark pair production are $\sigma_{tt} \leqslant 1.2$ pb [192] and $\sigma_{tt} \leqslant 89$ fb [193]. However, the latter limit is not fully model-independent, as the signal regions are optimized for higher mass vector mediators (with $M_V \geqslant 1$ TeV). As for our $Z'$ model, the highest possible $\sigma_{tt}$ we can

produce is 76 fb, for $M_{Z'} = 200$ GeV and $g_{ut} = 0.1$, which is just below the most stringent current experimental limits. $M_{Z'}$ is so low that we should expect the acceptance to be significantly affected by the experimental cuts. In particular, the $H_T \geqslant 750$ GeV cut imposed in Ref. [193] could be too high for such a low $M_{Z'}$. The cross-section $\sigma_{tt}$ is proportional to $(g_{ut})^4$ and decreases as we increase the mass, e.g. $\sigma_{tt} = 15$ fb for $M_{Z'} = 400$ GeV and $g_{ut} = 0.1$. All things considered, we conclude that prompt same sign top-quark pair production does not place relevant constraints on our parameter space.

Ref. [193] also casts experimental limits on ttj, although they consider a heavier Z′. Taking into account the re-scaling due to the $BR(Z' \to tj)$ and the same caveats about the acceptances, we find that these limits are also avoided by our model. A dedicated search for a light Z′ in this channel would be interesting, as one could potentially reconstruct the Z′ mass. Relaxing the $H_T$ cut and selecting events with a hard light-jet would potentially augment the signal acceptance, while the background would still be relatively small. Low-mass tj resonances have been searched in the $t\bar{t}+$ jets channel [194, 195], albeit with very low sensitivity.

While Z′-mediated single top-quark production $tj + \bar{t}j$ is absent due to $g_{uu}, g_{cc} = 0$, there is the possibility of Z′ radiative production tZ′j, where Z′ is emitted from either the $t/\bar{t}$ or the jet (which can then be either u, $\bar{u}$, c or $\bar{c}$). However, in our model this channel is also severely suppressed due to the chiral nature of our coupling choice. All radiative tZ′j production requires a $W^{\pm}$, either in the s-channel or the t-channel, with the latter dominating[2]. To interact with this $W^{\pm}$, we need left-handed quarks. However, in our model we require right-handed quarks to radiate a Z′, which yields a considerable suppression. In particular, due to pdf imbalance the most relevant diagram with $W^{\pm}$ for tZ′j is $ud \to du \to dtZ'$. But this diagram requires a left-handed u quark to interact with the $W$ and a right-handed u quark to interact with the Z′. As the u quark is essentially massless, the most important contribution to tZ′j essentially vanishes for right-handed couplings only. To illustrate this point, assuming $M_{Z'} = 400$ GeV, we find that the tZ′j cross-section for $g^L_{ut} = 0.0, g^R_{ut} = 1.0$ is approximately 40 times smaller than for $g^L_{ut} = 1.0, g^R_{ut} = 0.0$.

## 4.4 GLOBAL SEARCH AND KINEMATIC FEATURES

In Section 4.3 we show how a specific $Z^{'}$ vector boson can affect $t\bar{t}H$ and 4-top analyses. Although all BSM diagrams affect both observables to some extent, the charge asymmetric contribution mimics mainly $t\bar{t}W^{\pm}$ whereas the charge symmetric contribution mimics mostly 4-top behaviour in different searches and channels. In this Section we show how a more global study of the $Z^{'}$ could help to break this degeneracy and disentangle signal from SM background processes. Then, based on the studied discriminating observables, we define signal-enriched regions that have either $t\bar{t}W^{\pm}$ or 4-top as the main SM background and thus are sensitive to different parameters of the model. With the help of a few kinematical variables in these regions we aim to improve signal vs background discrimination, showing the potential of a future optimised Multivariate Analysis (MVA) discriminant.

2 A similar scenario is explored for a scalar in Ref. [196].

### 4.4.1 *Expected sensitivity of the global analysis*

The event selection used for the global analysis is similar to that of the 4-top search [25], but with some modifications. We define the regions as:

2LSS: Two same sign leptons, at least 3 jets and at least 1 b-jet,
3L: (Exactly) Three leptons, at least 2 jets and at least 1 b-jet.

No $H_T$ cut is applied in either of the regions.

The event categories in the global analysis are defined for the 2LSS and 3L selection separately, based on the discriminating variables: number of jets ($N_j$), number of b-jets ($N_b$), and total leptonic charge ($Q_{lep}$). In Fig. 25 we show the distributions for these discriminating variables in each category.

In order to perform a global analysis we define bins as $(N_j, N_b, Q_{lep})$ with $Q_{lep} = \pm$ as a shorthand for $\pm 2(1)$ for 2LSS (3L) and $\geqslant N_j$ ($N_b$) meaning at least $N_j$ jets ($N_b$ b-jets). There are in total 34 (30) bins for the 2LSS (3L) channel. Low $N_j$ and $N_b$ bins will have a larger contamination of $t\bar{t}W^{\pm}$ and higher $N_j$ and $N_b$ bins will be more sensitive to 4-top. The expected events for each one of these categories in the global analysis are shown in Fig. 26, where we use as a signal benchmark the combined $t\bar{t}H$ and 4-top best-fit point for $M_{Z'} = 400$ GeV from Fig. 21. We also plot the main irreducible backgrounds simulated at NLO accuracy with the same generator settings as for the signal, with cross-sections scaled to those reported in Ref. [25]. From the global picture we see how the signal mimics $t\bar{t}W^{\pm}$ in low $N_j$, low $N_b$ bins and 4-top in high $N_j$, high $N_b$ bins. There are some bins that provide a clear distinction between signal and background. More insight into how $t\bar{t}Z'$ can resemble SM 4-top is shown in Fig. 27, where both SM and BSM processes have a similar $H_T$ kinematic distribution in the events selected for the global histogram.

The global analysis shown in Fig. 26 can be used to estimate a potential discovery or the exclusion limit that can be set for the BSM $Z'$ model. This is done for $M_{Z'} = 400$ GeV and each point in the $(g_{ut}, g_{ct}, g_{tt})$ parameter space. Using Eq. 59 as detailed in Section 4.3, we generate global histograms for an arbitrary point in the parameter space using a fixed set of simulated points. These processes are not independent of each other and thus need to be considered simultaneously. An Asimov dataset [148] is used to calculate the expected significance and exclusion limits. The systematic uncertainties included in the global fit consist of 20% overall normalization uncertainty on the $t\bar{t}H$, $t\bar{t}W$, $t\bar{t}Z$, 4-top processes, and signal, as well as $N_j$- and $N_b$-dependent uncertainties on $t\bar{t}H$, $t\bar{t}W$, and $t\bar{t}Z$ to account for larger modelling uncertainties in the production of additional light/heavy-flavour-jets. We show in Fig. 28 the expected significance and the expected exclusion limits for the case where $M_{Z'} = 400$ GeV and $g_{ct} = 0$ for the corresponding integrated luminosity of the Run-2 dataset (139 $fb^{-1}$) and of the expected Run-2 plus Run-3 datasets (300 $fb^{-1}$).

### 4.4.2 *Disentangling signal and background in signal-enriched regions*

In order to define signal-enriched regions both for $tZ'$ and $t\bar{t}Z'$ we require $N_b \geqslant 3$. Moreover, given the top-quark content in each of these processes we can use $Q_{lep}$ and $N_j$ to further separate them. We therefore define two signal-enriched regions:

Region I, 2LSS (3L): $N_j \geqslant 4(3)$, $N_b \geqslant 3$ and $Q_{lep} = +2(+1)$,
Region II, 2LSS (3L): $N_j \geqslant 7(5)$, $N_b \geqslant 3$ and $Q_{lep} = \pm 2(\pm 1)$.

Region I is more sensitive to $tZ'$ while Region II is more sensitive to $t\bar{t}Z'$. We show the composition of the events that fulfil the selection of each Region in Fig. 29. We aim to examine observables that can disentangle signal from background in each one of the signal-enriched regions.

In the following, we plot only the fraction of events for the relevant processes in each region, i.e. $t\bar{t}W$ and $tZ'$ for Region I and 4-top and $t\bar{t}Z'$ for Region II. We show for $Z'$ the relevant masses 400 GeV and 600 GeV. We also find compelling to include in the study a BSM model in which we replace the spin 1 $Z'$ with a spin 0 scalar field $\Phi$ with a Lagrangian analogous to the Lagrangian in Eq. 58. For this scalar BSM model we use a mass $M_\Phi = 400$ GeV and the same couplings as the best-fit point of $Z'$ for this mass. The two relevant processes in this model are then $t\Phi$ and $t\bar{t}\Phi$.

To compare the relevant studied processes we calculate the separation between processes, defined as [197]:

$$\text{Separation} = \langle S^2 \rangle = \frac{1}{2} \sum_{i=1}^{N_{bins}} \frac{\left(f_i^{sig} - f_i^{bkg}\right)^2}{f_i^{sig} + f_i^{bkg}}, \tag{73}$$

where $f_i^{sig}$ ($f_i^{bkg}$) is the signal (background) fraction of events in bin $i$.

We plot in Fig. 30 the $H_T$ distribution for both regions. We can see in both regions how the low mass (400 GeV) BSM scenarios mimic the SM distribution regardless of their spin ($Z'$ or $\Phi$), whereas the 600 GeV case begins to show some deviation from the SM expected distribution.

The above results show that disentangling the 400 GeV BSM from the SM is challenging. Considering $tZ'$ and $t\bar{t}W$ events, we see that in the diagrams the lines connecting both same sign leptons have different Lorentz structure. We therefore explore the azimuthal angular separation between the leptons $\Delta\phi(\ell^\pm, \ell^\pm)$, which is a spin-correlation-sensitive observable. We observe in Fig. 31 that this observable can distinguish to some extent the different contributions in Region I, whereas no distinction occurs in Region II.

Another useful observation is that the leptons in $tZ'$ all come from top quarks, whereas in $t\bar{t}W$ there is one lepton coming from a prompt $W$ boson. In contrast to leptons coming from a top quark decaying to a $W$ boson and then to a lepton, the lepton coming from a prompt $W$ boson is not closely connected to a b-jet. Motivated by this, we propose a new kinematic variable defined as:

$$\begin{aligned} \text{MaxMin}(\ell, b) \quad = \quad & \text{The maximum of the minimum } \Delta R\text{-distances} \\ & \text{between the same sign leptons and a b-jet.} \end{aligned} \tag{74}$$

In the extreme case of boosted top quarks, we expect the $tZ'$ signal to have smaller MaxMin$(\ell, b)$ than the $t\bar{t}W$ background. For less boosted top quarks we expect this qualitative behaviour to still hold, although at a lesser extent. We show the MaxMin$(\ell, b)$ distribution for both regions in Fig. 32, where the expected behaviour is verified. Moreover, we find a slightly larger separation than for the $\Delta\phi(\ell^\pm, \ell^\pm)$ observable. We have verified that the MaxMin$(\ell, b)$ observable has better separation for larger $H_T$, as expected from its construction.

The previous observables, $\Delta\phi(\ell^\pm\ell^\pm)$ and MaxMin$(\ell, b)$, have shown a separation power between $tZ'/\Phi$ and $t\bar{t}W$. We are interested in understanding how independent is the separation power of these observables. We show therefore in Fig. 33 a two-dimensional event histogram where we plot the distribution over both observables to

distinguish between signal and background. To avoid low Monte Carlo statistic, we use the global analysis selection, without further splitting in Region I and II. There, we count events for all background processes and for the three different signal processes. We see from the figure that the two studied observables are not strongly correlated and can therefore be exploited simultaneously in a future MVA analysis.

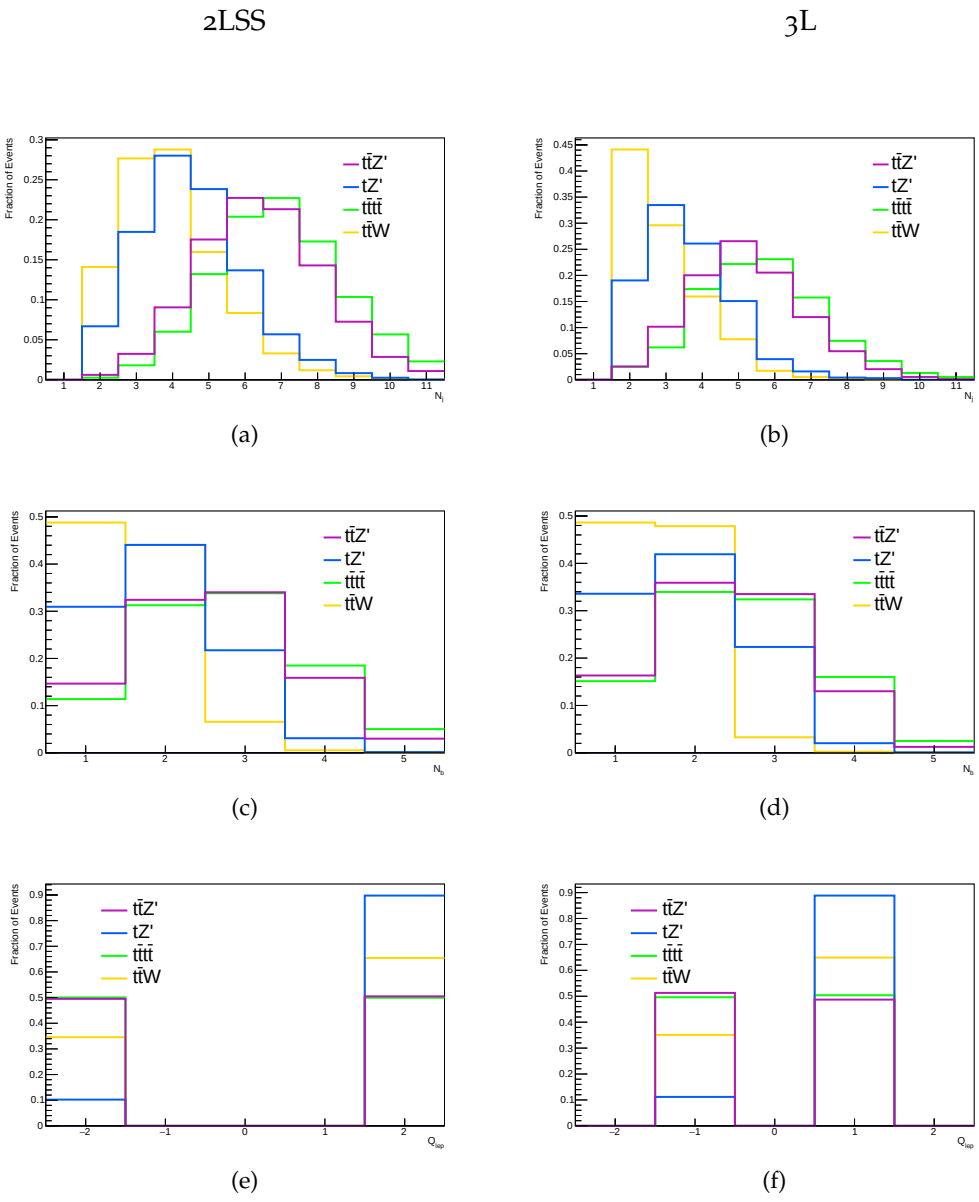
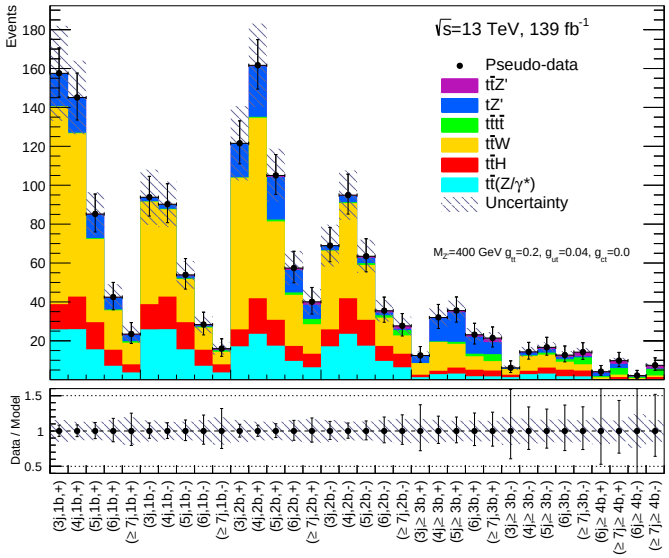
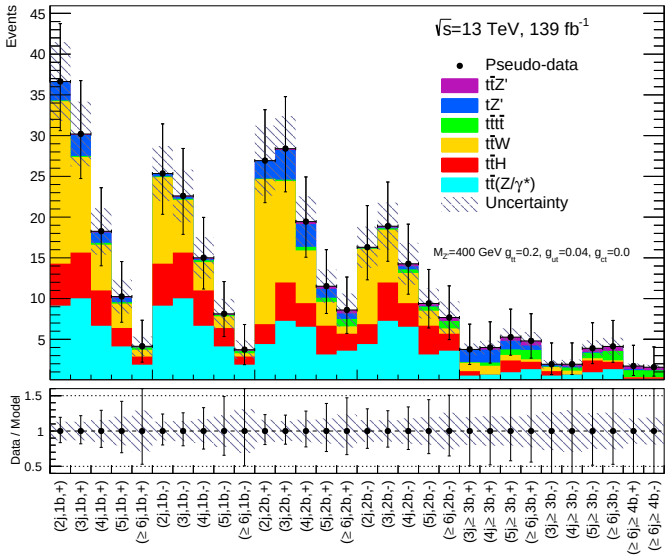



Figure 25: $N_j$, $N_b$, and $Q_{lep}$ distributions for the 2LSS (left) and 3L (right) categories used for the global analysis. Observe how $tZ'$ mimics $t\bar{t}W$ and $t\bar{t}Z'$ mimics $t\bar{t}t\bar{t}$, particularly in the $N_j$ distribution.




2LSS




3L

Figure 26: Global event yields for $tZ'$, $t\bar{t}Z'$ and the main irreducible backgrounds for the benchmark point $M_{Z'} = 400$ GeV, $g_{ut} = 0.04$, $g_{ct} = 0.0$ and $g_{tt} = 0.2$. Each bin consist of a global selection criteria on $N_j$, $N_b$, and $Q_{lep}$, as explained in the text.

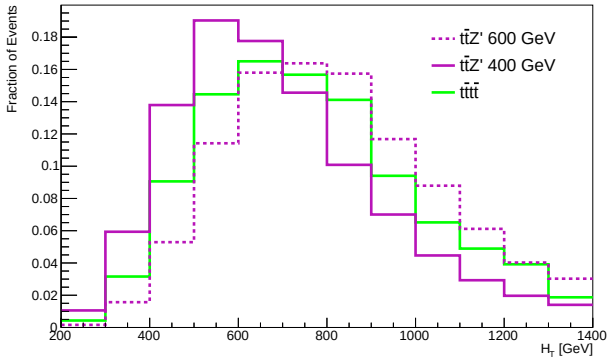
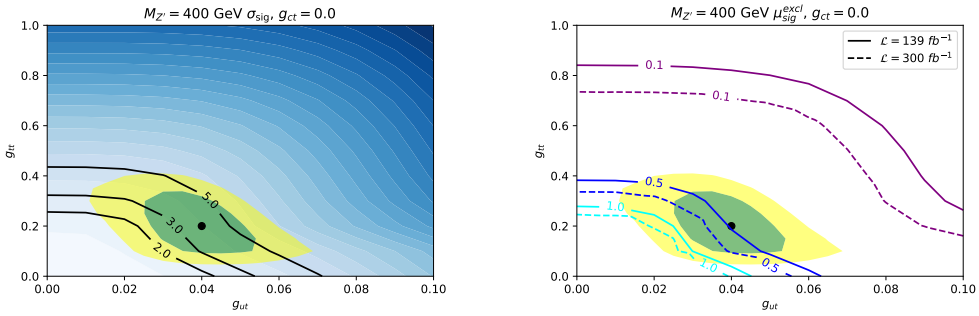
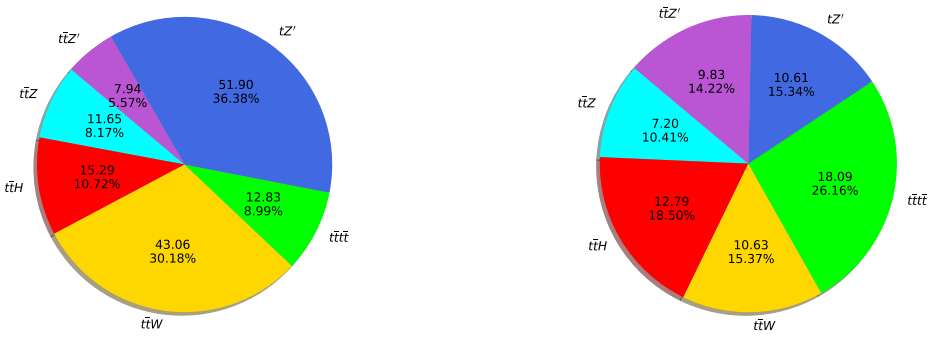



Figure 27: $H_T$ kinematic distribution for events belonging to the global analysis selection.



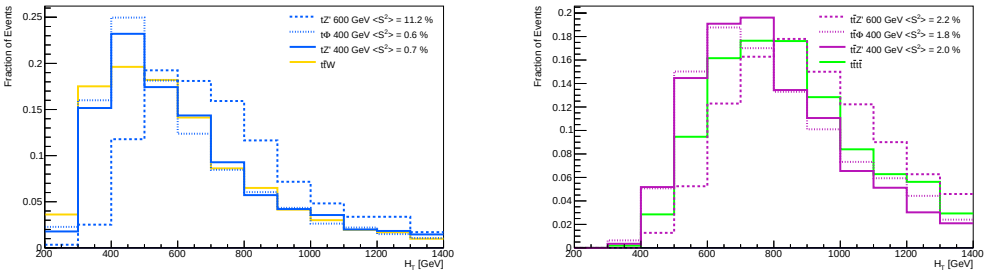

(a) (b)

Figure 28: Left: Expected significance using global bins for 139 $fb^{-1}$. Right: Expected exclusion limits on the signal strength μ. Solid (dashed) lines are limits for 139 $fb^{-1}$ (300 $fb^{-1}$). We also show the corresponding 1 and 2 s.d. regions of Fig. 21b.



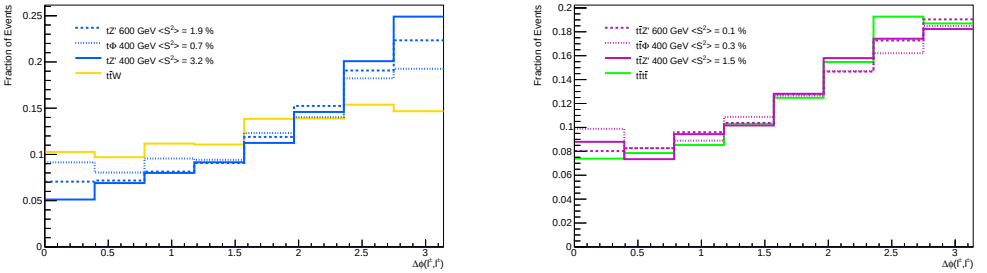

Region I Region II

Figure 29: Event yields and composition of Regions I and II.



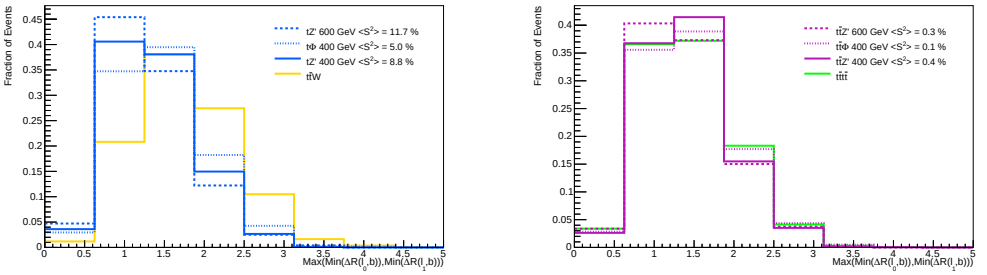

Region I Region II

Figure 30: $H_T$ fraction of events for both signal-enriched regions. As it can be seen, the 600 GeV BSM $H_T$ distribution has a separation an order of magnitude larger than the 400 GeV BSM scenarios for Region I, whereas for Region II all BSM scenarios are hardly distinguishable from SM.



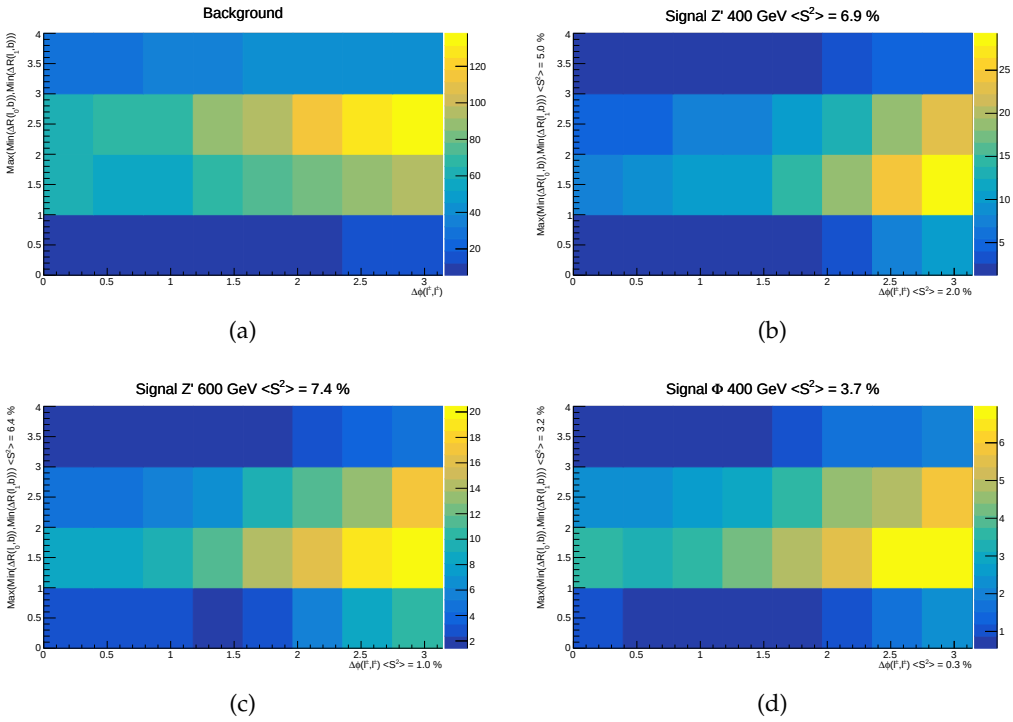

Region I Region II

Figure 31: Fraction of events of the azimuthal angular distance between the same sign leptons for the main processes in Regions I and II. We find it useful to disentangle $tZ'/\Phi$ from $t\bar{t}W$ in Region I.




Region I Region II

Figure 32: Fraction of events of the MaxMin$(\ell, b)$ observable. We find this observable to be useful for Region I which is dominated by $tZ'/\Phi$ and $t\bar{t}W$, since the latter has a lepton coming from a prompt $W$ boson.




(a) (b)

(c) (d)

Figure 33: Simulated event yields for different processes in the two dimensional space formed by the azimuthal angular distance between the two same sign leptons (x-axis) and the MaxMin$(\ell, b)$ observable (y-axis). When comparing the reported separations to those in Figs. 31 and 32, one should take into account that this is for the global selection, not Region I.

## 4.5 SUMMARY AND OUTLOOK

In this Chapter we have addressed the pattern of mild but persistent anomalies in same sign dilepton and multilepton plus b-jets events at the LHC from a phenomenological point of view. More precisely, we have studied a BSM model with a $Z'$ in a mass range 200 GeV to 600 GeV that couples hierarchically to the right-handed up-type quarks $u_R$, $c_R$ and $t_R$. The hierarchy differentiates the third generation from the other two, allowing for $t\bar{t}$ diagonal couplings and FCNC $t\bar{u}$ and $t\bar{c}$ couplings. We have shown how this model has the key features needed to reproduce the observed discrepancies in $t\bar{t}H$ and 4-top analyses.

Throughout this Chapter we focus primarily on the ATLAS analysis for $t\bar{t}H$ in Ref. [151] in which the post-fit yields of $t\bar{t}W^{\pm}$ show an excess over the expected cross-section by a factor of $\mu_{t\bar{t}W} = 1.39^{+0.17}_{-0.16}$. We study the case in which the $t\bar{t}W^{\pm}$ cross-section is not free-floating in the fit, allowing the BSM model to fill this excess. We explore this scenario by implementing the BSM model, simulating relevant processes up to detector level following the ATLAS analysis in two lepton same sign (2LSS) and trilepton (3L) final states, and finding the BSM parameters that best fit the ATLAS results. We perform a new simultaneous fit to the observed ATLAS data in the eight bins corresponding to combinations of dilepton charge and b-jet multiplicity in 2LSS and 3L final states (Fig. 2 in Ref. [151]). The regions in parameter space that best fits the ATLAS data are shown in Figs. 20-22. We show in Fig. 23 how the BSM model improves the interpretation of the ATLAS results in Ref. [151] without the need of re-scaling the $t\bar{t}W^{\pm}$ cross-section. These results indicate that for this observable the data would prefer over the SM a $Z'$ with mass 400 GeV to 600 GeV and couplings in the order $g_{tt} \sim \mathcal{O}(10^{-1})$, $g_{tu} \sim g_{tc} \sim \mathcal{O}(10^{-2})$. When the BSM model is also tested to reproduce the slight excess in the ATLAS 4-top analysis in Ref. [25], the preferred region for $g_{tt}$ is reduced to approximately half its value, since otherwise the model would populate this analysis with too many events.

We have verified that the best-fit regions are safe to other observables such as D-meson mixing, $Z'$-induced top quark rare decays, $t\bar{t}$ production, same sign top-quark pair searches, $tj$ resonance searches in $t\bar{t}j$ events, and $Z'$-mediated single top-quark production.

In the second part of this Chapter we explore how the proposed BSM model could be distinguished from similar SM physics processes in the multilepton plus b-jets final state. In a first stage we define 34 (30) kinematic regions for the 2LSS (3L) channel that have different signal-to-background ratios and therefore could be exploited to disentangle them. We perform a simultaneous fit of the regions in this global analysis search to the Asimov dataset assuming 139 $fb^{-1}$ and 300 $fb^{-1}$ integrated luminosity, and determine the prospects for discovering or excluding the proposed BSM model (Fig. 28). In a second stage, we define two kinematic regions tailored to increase the fraction of $tZ'$ and $t\bar{t}Z'$ events, respectively, and study observables that could help discriminating signal from the main SM background processes in each region. We find that $H_T$ could potentially discriminate $tZ'$ and $t\bar{t}Z'$ for $M_{Z'}$ above $\sim$ 600 GeV, whereas the $H_T$ distribution of $t\bar{t}Z'$ is very similar to that of the SM 4-top process. We also find that the azimuthal separation between same sign leptons $\Delta\varphi(\ell^{\pm}, \ell^{\pm})$ is a good variable to separate signal from its main background. In all cases, we include the possibility of replacing the $Z'$ with a scalar $\Phi$ field and the same flavour structure, and find that the $\Delta\varphi(\ell^{\pm}, \ell^{\pm})$ observable has discriminating power to differentiate the three models: SM, $Z'$ and $\Phi$. Finally, we propose a new kinematic variable $\mathrm{MaxMin}(\ell, b)$

(c.f. Eq. 74), which is likely to have a larger value if there is a lepton coming from a prompt $W$ boson in comparison to a $W$ boson coming from a top quark, which has a close b-jet as the top quark is boosted. This variable has a good separation power between signal and background. We find that it is easier to distinguish signal from background in Region I, in which $tZ'$ is the main signal. As a last test, we investigate for all events in the previous global analysis whether $\Delta\varphi(\ell^{\pm}, \ell^{\pm})$ and $\mathrm{MaxMin}(\ell, b)$ are correlated by plotting their distribution in a 2D coloured histogram, and find them to have little correlation and thus considering them together improves the separation power in comparison to each observable on its own.

In summary, we have proposed a phenomenological FCNC $Z'$ model that couples hierarchically to the up-type right-handed quarks to explain LHC discrepancies in same sign dilepton and multilepton plus b-jet final states. We have found regions in parameter space that fit the data better than the SM and proposed different ways to explore the data to test the BSM $Z'$ model. We find that a sophisticated experimental search, along the lines of our proposed global analysis could in the near future shed light on the existence of such a BSM scenario.

# Part II

# MACHINE LEARNING AT THE LHC

# 5

# 4-TOP DEMIXING

In this Chapter we apply an algorithm previously implemented for quark/gluon tagging in Ref. [82] to the problem of disentangling 4-top production from two of its main backgrounds in the same sign dilepton channel, $t\bar{t}H$ and $t\bar{t}W$. We take this as an opportunity to test the performance of the algorithm when the ideal hypotheses which underlie it are not entirely valid. This Chapter is based on work done in collaboration with Ezequiel Alvarez and Federico Lamagna in Ref. [198]. All necessary code to reproduce our results is available at Ref. [199].

As detailed in Subsection 1.1.1, 4-top production is one of the last benchmarks of SM Physics and could be specially sensitive to BSM effects [27–34]. In practice, measuring the 4-top cross-section is very challenging. A very populated final state and many different decay channels render its reconstruction very inefficient. 4-top searches [26, 200] rely on signal regions which are selected and analyzed with the help of Monte Carlo predictions. Although Monte Carlo predictions have reached a complex level of development, including Next to Leading Order (NLO) matrix level calculations for 4-top production [201], there are still many measurements of 4-top signal and backgrounds that require re-scaling to data in control regions which are in need of further understanding. Because of this, we investigate a direction to reduce the impact of Monte Carlo simulations and tuning in the extraction of physical quantities in the 4-top final state. It is worth stressing at this point that, even with all the new Machine Learning techniques at our disposal, it is not possible to avoid a dependence on Monte Carlo simulations in extracting absolute physical quantities from the LHC data. We pursue the goal of reducing the impact of these simulations by replacing some of the required generation and/or calibrations with different techniques.

In particular we focus on a Topic Modeling technique which we refer to as the *Topic Model Demixer algorithm*, or *Demixer algorithm* for short [82]. This technique was originally designed as a data-driven implementation of a mixture membership topic model and has been applied to differentiate quark from gluon jets in Refs. [82, 202]. It is an algorithm suitable for the simple case in which there are two observed distributions from which two underlying distributions need to be recovered. Along this Chapter we investigate this particular framework applied to 4-top physics, where the 4-top final state and its backgrounds are two specific topics which are mixed in different proportions in two defined samples according to the number of jets that are b-tagged in the final state. We explore within this framework the perspectives for measuring 4-top physical properties, combining the information that can be extracted from the Demixer algorithm with Monte Carlo simulations and techniques.

This Chapter is divided as follows. In Section 5.1 we present the Demixer algorithm and investigate its behavior in cases where its hypotheses are not fully satisfied. In Section 5.2 we apply the Demixer algorithm to 4-top production and its main backgrounds $t\bar{t}W$ and $t\bar{t}H$, where we show how to recover important physical properties and distributions without relying on Monte Carlo simulations. We also essay Machine Learning techniques on the Demixer algorithm to improve the results in describing the signal and background distribution and purity fractions. In Section 5.3 we discuss improvements and alternative strategies that could be performed on the algorithm to

enhance the results presented in this Chapter. We discuss in Section 5.3 a possible strategy to tune a Monte Carlo generator using the information extracted from the Demixer algorithm. We summarize our conclusions in Section 5.4.

## 5.1 THE DEMIXER ALGORITHM

In this Section we give a brief review on the Demixer algorithm. We first describe the algorithm along with the basic features that allow to recover two underlying distributions and their fractions from a pair of mixed samples. We then discuss under which conditions the algorithm works properly and we analyze more realistic cases which relax some of these conditions. We study how the algorithm can still be used to recover sensible topics as long as the departure from the hypotheses is tractable.

### 5.1.1 *Demixer algorithm in the ideal case*

Along this Chapter we consider the case of two samples $M_1$ and $M_2$ that are statistical mixtures, in different proportions, of two underlying sources which we call signal and background. In the following paragraphs we summarize the basic layout of the problem for the ideal case, but further details can be found in Refs. [82, 203].

Assuming that some features of the elements in samples $M_{1,2}$ can be described by a given observable $x$, we can define a probability distribution of these elements in each sample as

$$P_{M_i}(x) = f_i\, P_S(x) + (1 - f_i)\, P_B(x). \tag{75}$$

Here $P_S(x)$ and $P_B(x)$ are the two underlying distributions in $x$ of signal and background, respectively, and $f_i$ is the fraction of signal events in sample $M_i$. These are the unknowns that one would like to recover, or *demix*, from the original samples $M_1$ and $M_2$. Observe that one of the important assumptions is that $P_S(x)$ and $P_B(x)$ are the same for both samples.

In order to demix these samples we perform a maximal subtraction through the definition of reducibility factors:

$$\kappa_{ij} \equiv \kappa\left(M_i|M_j\right) \equiv \min\left(\frac{P_{M_i}(x)}{P_{M_j}(x)}\right) \tag{76}$$

We maximally subtract the $M_2$ sample from the $M_1$ sample and normalize it in order to define the distribution of the reconstructed signal topic $T_S$ over $x$ as:

$$P_{T_S}(x) = \frac{P_{M_1}(x) - \kappa_{12}\, P_{M_2}(x)}{1 - \kappa_{12}}. \tag{77}$$

In a similar way, we can obtain the reconstructed background topic $T_B$ distribution in $x$. Without loss of generality, we suppose that sample $M_1$ has a larger fraction of signal events than sample $M_2$ ($f_1 > f_2$), and therefore $T_S$ is the one matching the signal distribution, whereas topic $T_B$ would match the background distribution.

For the Demixer to recover exactly $P_S$ and $P_B$ we need, in addition to having different fractions $f_i$ in Eq. 75, to have anchor bins $x_S$, $x_B$ such that

$$\begin{aligned} P_S(x_S) = 0, \quad P_B(x_S) \neq 0, \\ P_S(x_B) \neq 0, \quad P_B(x_B) = 0. \end{aligned}$$

Defining the irreducibility factors for the underlying distributions $\kappa(B|S)$ and $\kappa(S|B)$ by replacing $M_{i,j}$ by S and B in Eq. 76, we see that the presence of anchor bins implies $\kappa(B|S) = \kappa(S|B) = 0$. This hypothesis is usually called *mutual irreducibility*. When this condition is guaranteed, topics reconstructed by maximally subtracting samples will match underlying signal and background distributions, and the fractions $f_{1,2}$ can be recovered from inverting the following equations:

$$\kappa_{12} = \frac{1 - f_1}{1 - f_2}, \qquad \kappa_{21} = \frac{f_2}{f_1} \tag{78}$$

If there is no mutual irreducibility, demixing the samples still leads to relevant quantities. For instance, the topic obtained from the more signal-like sample leads to the background-subtracted signal distribution:

$$P_{T_S}(x) = P_{S|B}(x) = \frac{P_S(x) - \kappa_{SB}\, P_B(x)}{1 - \kappa_{SB}}. \tag{79}$$

And the analogous is valid for the other sample, by swapping B and S. If there is extra input on the size of these two (in general different) reducibility factors, either by theoretical principles or by some given estimation, then the two equations can be solved to obtain the pure distributions. In this sense, $\kappa_{SB}$ and $\kappa_{BS}$ can be thought of as hyperparameters, since prior information on them would provide a better determination of the underlying topic distributions.

### 5.1.2 *Demixer algorithm beyond the ideal case*

In the above paragraphs we have shown a method for extracting two topics from two samples that contain different mixtures of said topics in an ideal case. In the real case scenario the procedure is highly different and many factors may affect the conclusions. For example, the minimum in Eq. 76 may not correspond to the true minimum due to the data uncertainties; or one of the underlying distributions may not have an anchor bin; or the signal or backgrounds distributions may be different in $M_1$ and $M_2$. In the following paragraphs we perform a brief study on how some of these real case scenario factors may affect the extraction of topics and fractions. In this overview we neglect experimental factors such as statistic and systematic uncertainties and we focus in studying cases where some of the hypotheses needed for the Demixer algorithm procedure are not satisfied. A more exhaustive study scrutinizing the effect of all these factors lies beyond the scope of this Chapter. Nevertheless, it would be useful for further understanding the reliability of the Demixer algorithm in particle physics where not only the ideal case is always far from reality, but also because the topics (and their fractions) are not abstract distributions but instead are meaningful physical quantities.

Let us consider the cases in which the mixture samples stray from the optimal conditions for demixing. In order to keep track of these conditions, we define the two basic hypotheses as:

- **H1:** Mutual irreducibility. Both underlying signal and background have anchor bins.
- **H2:** Same underlying distributions. The two samples are sums of the same signal and background distributions, differing only on the fractions.

In many real case scenarios H1 is relaxed either because mutual irreducibility is a priori unknown, or because it is well known that one of the underlying distributions does not have an anchor bin. Relaxing H1 means that the reconstructed topics now match background-subtracted-signal distribution and signal-subtracted-background distribution, as in Eq (79). However, if one knows the reducibility factors $\kappa_{SB}$ and $\kappa_{BS}$, inverting this system of equations would yield the true underlying distributions. In such cases one option is to resort to either theoretical or experimental arguments, or simulations in order to estimate the value of the reducibility factors. In the same direction, one could justify that one of the reducibility factors is zero –or negligible– and extract useful information, even though the other is unknown. In fact, as shown in Ref. [202] or in Eq. 79, this means that one can reconstruct the distribution that has an anchor bin exactly without recurring to additional information.

In order to quantitatively discuss deviations from H2 we need to quantify the difference between probability distributions. There are many options to quantify the difference between two curves. We decide to use the Hellinger distance [204] which is defined as

$$d_H^2(P_A, P_B) \equiv \frac{1}{2}\int dx \left(\sqrt{P_A(x)} - \sqrt{P_B(x)}\right)^2 = 1 - \int dx \sqrt{P_A(x) P_B(x)}. \tag{80}$$

because it is a natural distance to use for probability distributions. This distance has the property that it provides a relative enhancement in importance to regions with smaller values through the use of a square root. Because of this, the Hellinger distance is more sensitive to differences in the small probability regions, which are the regions where the anchor bins are located. It is worth mentioning that we have verified that using other notions of distance such as $L^1$ or $L^2$ does not yield qualitative changes to our conclusions.

With this notion of distance, we can measure the Demixer performance when the hypothesis of same underlying distributions is not fulfilled. We define $\delta_{S,B}$ to quantify that either the signal or background distributions can be different in each sample:

$$\delta_S \equiv d_H(P_{S_1}, P_{S_2})$$
$$\delta_B \equiv d_H(P_{B_1}, P_{B_2}).$$

Here the sub-index 1 or 2 refers to the underlying distributions in samples $M_1$ and $M_2$, respectively. To determine the performance of the algorithm, we can measure how good signal and background reconstruction is by defining the distances

$$\Delta_{S_i} \equiv d_H(P_{T_S}, P_{S_i})$$
$$\Delta_{B_i} \equiv d_H(P_{T_B}, P_{B_i}).$$

Since sample $M_1$ is the one with larger fraction of signal, and sample $M_2$ the one with larger fraction of background, we define for practicality $\Delta_S = \Delta_{S_1}$ and $\Delta_B = \Delta_{B_2}$.

In order to deviate from the second hypothesis without deviating from the first, we keep the same notion of mutual irreducibility when in presence of four underlying samples $P_{S_{1,2}}(x)$ and $P_{B_{1,2}}(x)$ by modeling the deviation between underlying samples to be generated by multiplicative noise, that is

$$P_{S_2}(x) = P_{S_1}(x)\,(1 + \xi_S(x))$$

with a noise function $\xi_S$ satisfying

$$\int dx\, P_{S_1}(x)\, \xi_S(x) = 0 \tag{81}$$

to keep $P_{S_2}(x)$ normalized. For the background we have analogous relationships exchanging $S \to B$ and $1 \leftrightarrow 2$. $P_{S_1}(x)$ and $P_{B_2}(x)$ are thus the reference distributions for signal and background, respectively.

To better explore the Demixer algorithm conditions and their subsequent relaxation we split the space of possible combinations into a few cases. For H1 we have either

- Case a) Mutual irreducibility
- Case b) Only one anchor bin
- Case c) No anchor bins

If H2 holds, in case b) the algorithm still recovers correctly the distribution that has the anchor bin. If in addition the reducibility factors are known, in all three cases the Demixer works in recovering underlying distributions. For H2 we distinguish three cases:

- Case 0) $P_{S_1}(x) = P_{S_2}(x), \quad P_{B_1}(x) = P_{B_2}(x)$
- Case 1) $P_{S_1}(x) \neq P_{S_2}(x), \quad P_{B_1}(x) = P_{B_2}(x)$
- Case 2) $P_{S_1}(x) \neq P_{S_2}(x), \quad P_{B_1}(x) \neq P_{B_2}(x)$

Case 0) is where H2 holds. For case 1) there is a mirrored case that occurs by switching between S and B labels. Case 2) is the most general one, and can be particularized to the other cases for either $\delta_B \to 0, \delta_S \to 0$ or both.

We can see how the Demixer works when H2 does not hold. In this case the factors $\kappa_{ij}$ get modified due to the difference between underlying signal and background distributions

$$\kappa_{12} = \left(\frac{1-f_1}{1-f_2}\right)\frac{P_{B_1}(x_S)}{P_{B_2}(x_S)}, \quad \kappa_{21} = \left(\frac{f_2}{f_1}\right)\frac{P_{S_2}(x_B)}{P_{S_1}(x_B)}. \tag{82}$$

As we parametrize the difference between functions as a multiplicative noise, we get

$$\kappa_{12} = \left(\frac{1-f_1}{1-f_2}\right)\left(1+\xi_B^0\right), \quad \kappa_{21} = \left(\frac{f_2}{f_1}\right)\left(1+\xi_S^0\right). \tag{83}$$

Here $\xi_B^0$ and $\xi_S^0$ are the values of the noise functions at the anchor bins, which should not be confused with the functions $\xi_B(x)$ and $\xi_S(x)$. Performing the maximal subtraction on sample $M_1$ yields the signal-topic

$$\begin{aligned} P_{T_S}(x) &= P_{S_1}(x)\frac{(1-f_2)f_1 - f_2(1-f_1)(1+\xi_S(x))(1+\xi_B^0)}{1-f_2-(1-f_1)\xi_B^0} \\ &\quad + P_{B_2}(x)\frac{(1-f_1)(1-f_2)\left(\xi_B(x)-\xi_B^0\right)}{1-f_2-(1-f_1)\xi_B^0} \end{aligned} \tag{84}$$

The background-topic is found by replacing $f_1 \to 1-f_2$, $f_2 \to 1-f_1$, and flipping B and S labels. We can first consider the simpler case 1), which corresponds to $P_{B_1}(x) \equiv P_{B_2}(x)$. In this case the expressions simplify to

$$P_{T_S}(x) = P_{S_1}(x) - \frac{f_2(1-f_1)}{f_1-f_2}P_{S_1}\xi_S(x) \tag{85}$$

$$P_{T_B}(x) = P_B(x)\left(1+\frac{f_1 f_2 \xi_S^0}{f_1 - f_2(1+\xi_S^0)}\right) + P_{S_1}(x)\frac{f_1 f_2\left(\xi_S(x)-\xi_S^0\right)}{f_1-f_2(1+\xi_S^0)} \tag{86}$$

We can see that in general the signal reconstruction is better than the background reconstruction, as the expression involves only the underlying signal distribution, whereas the background topic involves both signal and background distributions in addition to denominators which may enhance the disagreement between $P_{T_B}(x)$ and $P_B(x)$. More quantitative statements can be made, by calculating the distances $\Delta_B, \Delta_S$ and $\delta_S$. For instance, by computing $\Delta_S$ using Eq. 85 and expanding in $\xi_S$ we obtain

$$\begin{aligned}\Delta_S &= d_H(P_{T_S}, P_{S_1}) \simeq \left[\frac{(1-f_1)f_2}{f_1-f_2}\right] \frac{1}{2\sqrt{2}} \sqrt{\int dx\, P_{S_1}(x)\xi_S^2(x)} \\ &= \left[\frac{(1-f_1)f_2}{f_1-f_2}\right] d_H(P_{S_1}, P_{S_2}). \\ &= \left[\frac{(1-f_1)f_2}{f_1-f_2}\right] \delta_S. \end{aligned} \tag{87}$$

We see that $\Delta_S$, a measure on the performance in the signal reconstruction, follows a linear dependence on the distance between the two signal distributions, which measures the breaking of condition H2. We also observe that the slope is simply a function of the signal fractions in the samples $f_1, f_2$, and that by increasing $f_1$, that is the amount of signal in sample $M_1$, this slope decreases. This reflects the fact that the purer the samples on a given distribution, the better the reconstruction of the corresponding topic.

To better visualize the above results, we have performed numerical simulations by scanning on different function and noise shapes. We have used Gaussians for the S and B distributions, randomly sampling their means and deviations and keeping only cases with mutual irreducibility under a certain tolerance. For the noise functions $\xi_{S,B}(x)$ we sampled randomly for each bin a value between $[-0.1, 0.1]$, then we rescaled the functions in order to satisfy the relation Eq. 81 (and the analogous for the $P_B(x)$ distribution). Then two mixture samples having fractions $f_1 = 0.45$, $f_2 = 0.22$ were generated by adding the underlying distributions.

Using these simulations we have computed the distances for each topic. Each simulated point corresponds to a pair of mutually irreducible functions $P_{S_1}(x)$ and $P_{B_2}(x)$, and two noise functions $\xi_S$, $\xi_B$ of the same amplitude. In Fig. 34 we plot of these distances for Case 1), which corresponds to setting $P_{B_2}(x) = P_{B_1}(x)$ and using a single noise function $\xi_S(x)$. We see the linear dependence in signal reconstruction with the signal difference $\delta_S$. We have verified a similar result for other distances such as $L^1$. This behavior is well explained by Eq. 87. These points can be linearly interpolated for different values of $f_1 > f_2$, to see that the slopes follow the predicted values. From Eqs. 85 and 86 one can see that the background topic is expected to be noisier. In fact, one can see from the right-hand side in Eq. 87 that for fixed $\delta_S$, the distance $\Delta_{S_1}$ is approximately constant. Whereas the same procedure for $\Delta_B$ is more involved due to the other factors present in Eq. 86 which yield stochastic noise. We also see in Fig. 34 that the Hellinger distance $\Delta_B$ is an order of magnitude larger than $\delta_S$, showing that the background reconstruction is sensitive to the distance between the underlying signal distributions $\delta_S$. We find less enhancement if the $L^1$ distance is used.

For Case 2), with different signal and background distributions in each sample $M_i$, reaching a closed form for the distances $\Delta_B$ and $\Delta_S$ is considerably more involved. We have computed numerically these distances using the same simulation scheme as in the previous case and plotted the results in Fig. 35. For easy visualization, we plot the distances as a function of $\delta_B$ in the left panel and as a function of $\delta_S$ in the right

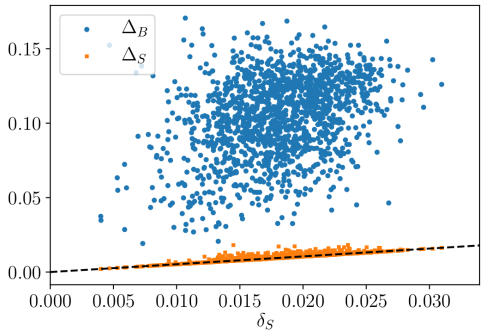



Figure 34: Testing H2 under case 1). Scatter plot of distance between the reconstructed signal and background topics and the corresponding underlying distribution as a function of distance between the two underlying signal distributions. We show in dashed black the line given by slope $(1-f_1)f_2/(f_1-f_2)$ from Eq. 87. The vertical axis corresponds to $\Delta_{S_1}$ for orange (square) points and to $\Delta_B$ for blue (circle) points.

panel. We extend the range of $\delta_B$ in the left panel to the region where $\delta_B \ll \delta_S$ and verify that we recover Fig. 34 in the limit $\delta_b \to 0$. In the right panel we only plot points where $\delta_S \sim \delta_B$. From using specific forms of $\xi(x)$ noise functions one can infer that the topic with more fraction in the samples is the one that would have a better reconstruction. We have verified this statement by simulating many cases as in Fig. 35 with different fractions. As we work with signals that are rarer than background even in the signal-enriched mixture, this means that we expect a better reconstruction of the background than of the signal.

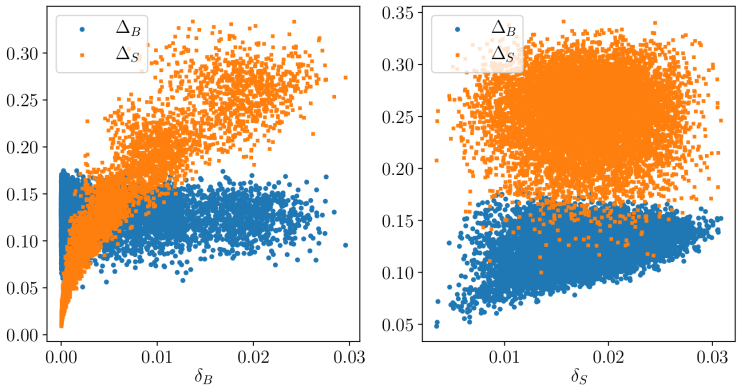



Figure 35: Testing H2 under Case 2. Scatter plot for $\Delta_B$ and $\Delta_S$ as functions of $\delta_B$ and $\delta_S$, for fractions $f_1 = 0.45$, $f_2 = 0.22$. In the left panel we have extended the simulations to also include values of $\delta_B$ well below $\sim 0.01$, in order to make the connection with Fig. 34 in the limit $\delta_b \to 0$. Vertical axis is as in Fig. 34.

From the above results we observe that there is an interplay between the distances of the underlying distributions and the fractions of purity of each topic in the samples to determine which topic is better reconstructed and to what extent. For instance, in Fig. 34 we see that having a non-zero distance between the signal distributions results in the background being less well reconstructed than the signal. This statement is still valid for specific values of the fractions $f_i$ that increase the slope of the orange points, reaching the limit in which both topics are equally reconstructed. On the other hand,

in Fig. 35 we see that for $\delta_S \sim \delta_B \neq 0$ the value of the fractions $f_i$ is the one that determines which topic is better reconstructed.

As a summary of the above paragraphs, we can extract some useful statements regarding the validity of the Demixer when the working conditions do not fully satisfy the H1 and H2 hypotheses.

- When mutual irreducibility is not guaranteed, if one of the topics has an anchor bin, then this topic is better reconstructed by the algorithm. (Case b0).
- If the underlying distributions for one topic are different in the two samples $M_i$, and the other topic has equal underlying distributions, then the former is the one with better reconstruction. (Case a1.)
- If both topics have different distributions over both samples, then topic reconstruction is mainly ruled by sample purities. If samples are mostly background, then background reconstruction will be better than signal reconstruction, and vice-versa. (Case a2.)

In the next Section we apply what we have learned to the case of 4-top production.

## 5.2 4-TOP PRODUCTION

The 4-top final state at the LHC is a very busy channel with a high degree of difficulty to reconstruct. In addition, this signal and its backgrounds suffer from important Monte Carlo uncertainties. It is therefore an attractive channel in which to apply the Demixer algorithm and where we can expect to find many of the difficulties discussed in Section 5.1.

To choose the final states in which to apply the Demixer algorithm, we note that the algorithm requires a non-negligible fraction of signal in each of the samples. Considering the top-quark Branching Ratios and background processes, we find it suitable to focus on the same-sign dilepton channel where the main backgrounds are $t\bar{t}W$ and $t\bar{t}H$ with the $W$ decaying leptonically and the $H$ decaying semi-leptonically or leptonically through $WW^*$. This channel provides the highest signal to background ratio, provided we take the appropriate cuts beforehand as in Ref. [26]. We have performed event Monte Carlo simulations of $pp \to t\bar{t}t\bar{t}$ up to detector level using `MadGraph5_aMC@NLO` [12] for matrix-level process, `Pythia 8` [16] for showering and hadronization and `Delphes 3`[18] for detector simulation, following the same basic cuts as in Ref. [26] but at leading order and with only up to one extra parton. We have set the mixture fractions to agree with the event yields reported in Ref. [26].

In order to have a good demixer we need to find two sets of observables as uncorrelated as possible, one to define $M_1$ and $M_2$ and the other one to play the demixing variable $x$ as defined in Section 5.1. Uncorrelated sets are needed to provide the approximately same underlying distributions for signal and background in both samples and comply with the H2 hypothesis. A naive set of observables from which to choose is $N_b$, $N_j$ and the $p_T$, energy, $\eta$ and $\phi$ from all the reconstructed objects in the event. In order to do a simple demixing model, we choose only one variable to define $M_1$ and $M_2$ and another one to demix on. This is detailed in Subsection 5.2.1. We also perform a multidimensional demixing with a Machine Learning algorithm encoding the multidimensional space into a single observable, what is called the CWoLA method [72], in Subsection 5.2.2 and we find that it does not show considerable improvements for

our current set-up.

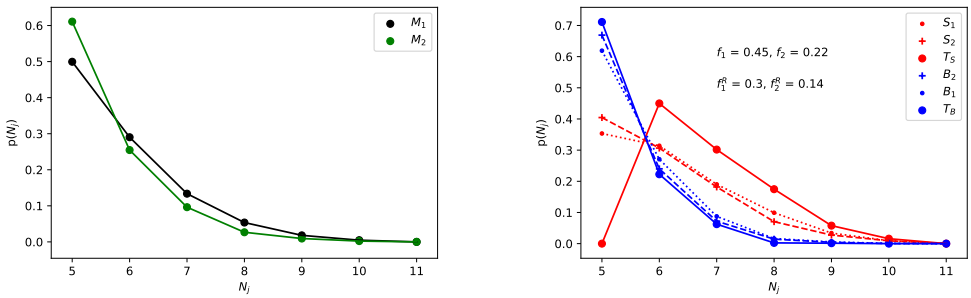
### 5.2.1 $N_j$ *demixing*

On physical grounds, the most direct distinction between signal and background would be the number of b-tags ($N_b$). Observe that the same sign dilepton channel has the special feature that none of the backgrounds has more than 2 b-jets. Meanwhile, and as showed in Ref. [28], the number of reconstructed jets ($N_j$) is also a good discriminator between signal and background. We consider these two variables to perform the $M_i$ definition and the demixing, leaving $p_T$, energy, $\eta$ and $\phi$ for setting cuts to accept or reject reconstructed objects such as leptons, jets or b-jets. To comply with the H2 hypothesis as much as possible we require the two background processes $t\bar{t}W$ and $t\bar{t}H$ contained in both $M_i$ to yield approximately the same underlying distribution on each sample. Since these backgrounds have different number of jets at parton level, it is suitable at this level to avoid dividing the samples using $N_j$ because the relative proportions of these backgrounds suffer a non-negligible change in each sample, yielding different underlying distributions for the background. In addition, $N_b$ is a variable with very few discrete values to use as a demixing observable. Therefore, we define the samples as

$$\begin{aligned} M_1: &\quad \text{Events with 3 b-tags,} \\ M_2: &\quad \text{Events with 2 b-tags.} \end{aligned}$$

whereas $N_j$ is the variable over which the demixing algorithm is applied.

Using the above $M_i$ definition we can construct the two mixture samples from the simulated events. We show in Fig. 36a the distribution of $M_i$ in $N_j$. As expected, the $M_1$ mixture –which has the largest fraction of signal– is shifted towards large $N_j$ with respect to the $M_2$ mixture. We have computed that the two distributions would be distinguishable at a $\sim 3\sigma$ level at the $N_j = 7$ bin with full LHC Run 3 luminosity.




(a) (b)

Figure 36: a) $M_1$ ($N_b = 3$) and $M_2$ ($N_b = 2$) distribution over $N_j$. b) Reconstructed underlying distributions using the Demixer algorithm. Solid lines represent the reconstructed topics after the Demixer algorithm is applied to the left panel samples. We also show in dotted and dashed the truth-level underlying distributions in each mixture sample.

We can apply the Demixer algorithm as described in Section 5.1 assuming, in a first step, that the data complies with the ideal case's two hypotheses H1 and H2. Since

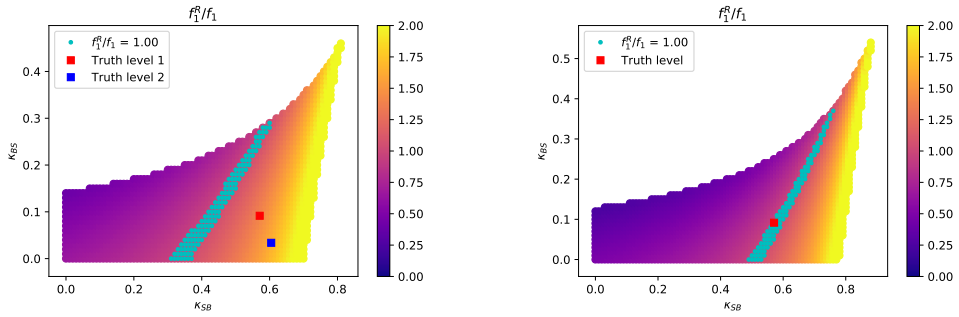
we can expect from general grounds that the background distribution goes to zero for large $N_j$ faster than the signal distribution, it is reasonable to expect a good background reconstruction. We plot the reconstructed underlying probability densities as well as the truth-level topic distributions in Fig. 36b. As expected, the background underlying distributions are approximated properly by the background reconstructed topic through the algorithm. As studied in Section 5.1, this is because the background has a proper anchor bin at large $N_j$, it has similar distributions in $M_1$ and $M_2$ and also because it is present with a larger fraction than the signal in both samples. Conversely, the reconstructed signal topic does not match the underlying signal distributions because of the lack of an anchor bin. As discussed below, obtaining a trustworthy background distribution in the signal region from data provides a new way of tuning the Monte Carlo event generator in the signal region.

Further analysis of the Demixer algorithm indicates that the reconstructed fractions misestimate their true values by $\sim 30\%$. This result is obtained through a Monte Carlo independent algorithm and the shift is within the order of magnitude of the usual Monte Carlo normalizations performed in 4-top signal and background predictions. Moreover, this result is the product of applying the Demixer algorithm assuming H1 holds, which we know is not true from theoretical grounds as the signal populates all $N_j$ bins. As a matter of fact, we can see in Fig. 36b that the algorithm yields a background-subtracted signal distribution which assigns a zero probability to $N_j = 5$. Therefore, we can still improve this result at the price of including Monte Carlo inputs.

As a second step, we address the H1 violation by using the defined workaround for the absence of anchor bins: the $\kappa$ factors defined in Section 5.1. These $\kappa$ factors can be understood as hyperparameters in the algorithm that model our prior understanding of the underlying topics. We can study the performance in the reconstructed fractions using the hyperparameter plane ($\kappa_{SB}$, $\kappa_{BS}$). In Fig. 37 we scan on these $\kappa$ for the real case of different underlying distributions in both samples (Fig. 37a) and for the adjusted case of equal underlying distributions in both samples (Fig. 37b). In both cases we see that using the prior theoretical knowledge that $\kappa_{SB} > 0$, and tuning manually $\kappa_{SB}$ to larger values while leaving $\kappa_{BS} = 0$, pushes $f_1^R/f_1$ to one. Moreover, we can see that when the underlying distributions are not equal in both samples (Fig. 37a), the $\kappa$'s that correctly reconstruct $f_1^R/f_1$ do not coincide with the solution corresponding to the correct underlying distributions. This is reflected in Fig. 38 where we show how by manually tuning $\kappa_{SB}$ to larger values we reach a solution in which the fractions are correctly reconstructed, but not the distributions, and then vice-versa. This behavior is still obtained if one also varies $\kappa_{BS}$, since the reason behind this disagreement is that H2 is not fulfilled.

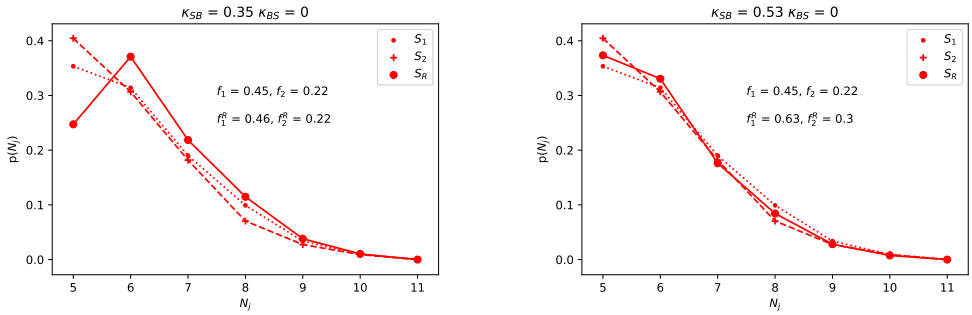
If, on the other hand, we satisfy H2 by forcing to have same underlying distributions (Fig. 37b), we can tune the hyperparameters $\kappa_{SB,BS}$ to correctly reconstruct the fractions and the distributions. For the sake of completeness, we show in Fig. 39 the output for the demixing algorithm in this case. The plots in Fig. 37 show the sensitivity of the algorithm to H2, as noticed in Section 5.1. We also see in this figure that the correct solution is more sensitive to $\kappa_{SB}$ than to $\kappa_{BS}$. This is expected, since the background does have an (approximate) anchor bin, and thus $\kappa_{BS}$ is expected to be close to zero.

In Section 5.3 we complete this discussion with possible methods to combine the Demixer algorithm with Monte Carlo simulations to reduce the impact of Monte Carlo tuning in the extraction of physical quantities from observations while obtaining accurate estimates of the necessary $\kappa$ factors.




(a) (b)

Figure 37: Reconstructed $f_1^R$ modified by $\kappa_{SB}$, $\kappa_{BS}$ in terms of the true fractions for a) different underlying distributions and b) same underlying distributions. The truth-level ($\kappa_{SB}$,$\kappa_{BS}$) pair are estimated using the underlying distributions in $M_1$ and $M_2$.




(a) (b)

Figure 38: Truth-level signal distributions, $S_1$ and $S_2$, and the background-subtracted signal topic distribution $S_R$ obtained using Eq. 79 for different values of $\kappa_{SB}$. Manually increasing $\kappa_{SB}$ while leaving $\kappa_{BS} = 0$ yields at some point the correct fractions but with disagreement of the $S_R$ and $S_1$ distributions (left panel), whereas further increasing $\kappa_{SB}$ yields correct distributions, but incorrect fractions (right panel). This is an explicit and graphical demonstration on the sensitivity of the algorithm to H2.

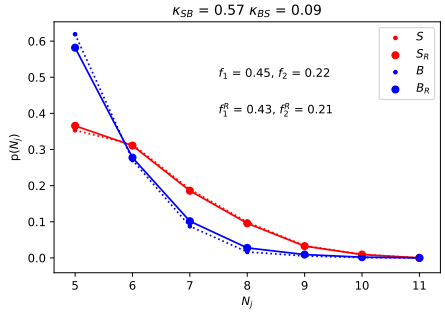



Figure 39: Corrected distribution and fractions for the truth-level ($\kappa_{SB}$, $\kappa_{BS}$) when forcing the underlying distributions to be the same. Here $S_R$ and $B_R$ stand for the distributions obtained using Eq. 79.

### 5.2.2 *CWoLA*

In Subsection 5.2.1 we detail an implementation of the Demixer algorithm to demix the 4-top signal probability distribution from two backgrounds in the same sign dilepton channel. The Demixer algorithm, as detailed in Section 5.1, is implemented in a simple and clear way with only one observable ($N_b$) to define the orthogonal regions $M_1$ and $M_2$ and one observable ($N_j$) to demix over and obtain the topics $T_S$ and $T_B$. In principle this could be improved as stated in CWoLA [72] with the use of a larger set of observables that have to be combined in some way to get an optimal classifier for $M_1$ and $M_2$ which corresponds to the optimal classifier for signal and background. In this sense, the implementation detailed in Subsection 5.2.1 can be regarded as the naive benchmark against which more sophisticated algorithms should be compared to see if the increase of complexity is warranted from a performance point of view. One should keep in mind, however, that to identify the output of CWoLA to the real signal and background distributions one still has to consider the validity of H1 and H2 as enunciated in Section 5.1.

In this Subsection we study the use of ML algorithms to search for a better discriminant than using only $N_j$, while we maintain the definition for $M_1$ and $M_2$ as a function of the number of $N_b$. This discriminant should provide a better resolution between $M_1$ and $M_2$ while also providing anchor bins for both the reconstructed distributions. These distributions should be identified with signal and background if the hypotheses in Section 5.1 are fulfilled.

As we are aiming to obtain an approximation to the probability densities of a multidimensional feature space, a sensible choice would be to use a Neural Network (NN). We implement a simple Fully Connected Feed Forward Deep Neural Network with the `Keras` package [205] for Python 3, which we will refer to simply as an NN. We feed the NN with the same simulated events, and label the events according to their classification into $M_1$ or $M_2$. We test different NN architectures and different observables of the reconstructed events. We split the samples into training, validation and testing samples. The training is performed using 200 epochs provided there is no overfitting. We make use of the classes weights option provided by `Keras` to account for the different number of events for $M_1$ and $M_2$ reported in Ref. [26] without discarding any simulated event. From each event we extract $N_j$, the $p_T$ and energy of all reconstructed objects, angular distance between any pair of reconstructed objects, and total transverse energy $H_T$. In the following we use layers of neurons with `ReLu` activation functions and a final neuron with a sigmoid activation function forming a feed-forward NN trained with a binary cross-entropy loss function. The notation for the chosen observables is $N_1$object$_1$-$N_2$object$_2$ which means that we use the $p_T$ and energy of the $p_T$-leading $N_1$ and $N_2$ objects object$_1$ and object$_2$ and the $\frac{1}{2}(N_1+N_2)(N_1+N_2-1)$ angular distances between all of them. For instance, in "Nj-2leps-2jets" we refer that the NN is being fed with $N_j$ and all the information of the 2 leading leptons and 2 leading jets in form of $p_T$, energy and angular distances between them.

Along Figs. 40, 41 and 42 we show three simple architectures with three different choices of observables in each figure. We plot the NN output for $M_1$ and $M_2$ and the result of using the Demixer algorithm on this output to recover the underlying distributions and the signal fractions in each mixture. As a general result, we find that using the b-jets when feeding the NN brings correlation between the demixing observable and the sample definition which causes the algorithm to fail to reconstruct the underlying distribution even though the discrimination is efficient. This is ex-

plained by the fact that we violate H2 hypothesis. If we instead feed the NN with the leading jets –regardless of whether they are b-tagged or not– then such a correlation is suppressed and the reconstruction of the underlying distributions is better. We also find that using less jets proves to be slightly better. In Fig. 40 we investigate NN with 3 layers of 4 nodes each and a final sigmoid node, which we refer to it as "4+4+4+1". We find that the NN is too simple and the reconstructed and underlying topics consist of tacking lines. In Fig. 41 we add nodes to each layer and use a 32+32+32+1 NN to obtain smoother curves for reconstructed and underlying topics. Adding one more layer with 32 nodes brings still more smoothness and a good agreement in the reconstructed fractions as shown in Fig. 42.

We find the best set up by using a 32+32+32+32+1 NN fed with $N_j$, the two leading leptons and the two leading jets, Fig. 42c and Fig. 42d. These results should be compared to Fig. 36. We see that the improvement in performance is not enough to justify the added complexity of the NN. This lack of significant improvements can be understood by the fact that $N_j$ is by far the most important feature among the information we feed the NN, and it was already present in Subsection 5.2.1. Any future work on improving these results will thus need to a data representation which improves significantly on the discriminating power of using $N_j$ alone.

## 5.3 DISCUSSION AND FUTURE

Along previous Sections we have studied how the Demixer algorithm works in recovering signal and background distributions from mixture samples, and how to apply these tools to the 4-top process. In this Section we discuss the strengths and some of the shortcomings of this implementation, along with possible improvements. We first discuss possible goals relative to the analysis in Section 5.1, when the Demixer hypotheses are not fully satisfied. We then discuss possible alterations to the Demixer algorithm implementation presented in Section 5.2 together with its pros, cons, and features to be further explored. We end with a discussion on how the presented Demixer algorithm could be implemented to extract physical quantities in 4-top while reducing the impact of Monte Carlo simulations.

In Section 5.1 we detailed the hypotheses behind the topics and fractions extraction, exploring what happens when these hypotheses are not valid. There is still many aspects to explore, both analytically and numerically. For instance, the error propagation of the Demixer algorithm can be studied for the more realistic case in which the uncertainty in each sample is taken into account. Error bars in samples $P_{M_1}(x)$ and $P_{M_2}(x)$ would be translated into error bars in $P_S(x)$, $P_B(x)$, along with reconstructed fractions $f^R_{1,2}$. Analyzing this error propagation and its dependence on the validity of conditions H1 and H2 defined in Section 5.1 would be an important step towards better understanding the algorithm and its potential.

In Section 5.2 we considered how to apply the Demixer algorithm to the problem of extracting the 4-top signal distribution in the same sign dilepton channel. The main reasons for choosing the same sign dilepton final state are the relatively high S/B ratio, and also that this channel has the special condition that all background processes consist of at most two b-jets at parton level. It is worth noticing at this point that other minor backgrounds, such as non-prompt leptons and $t\bar{t}Z$, have a similar behavior to the main backgrounds in what concerns to the number of b-jets and to the anchor bin for large $N_j$. We have also studied other channels such as monolepton, for which we only mention the following results. We find that the S/B ratio,

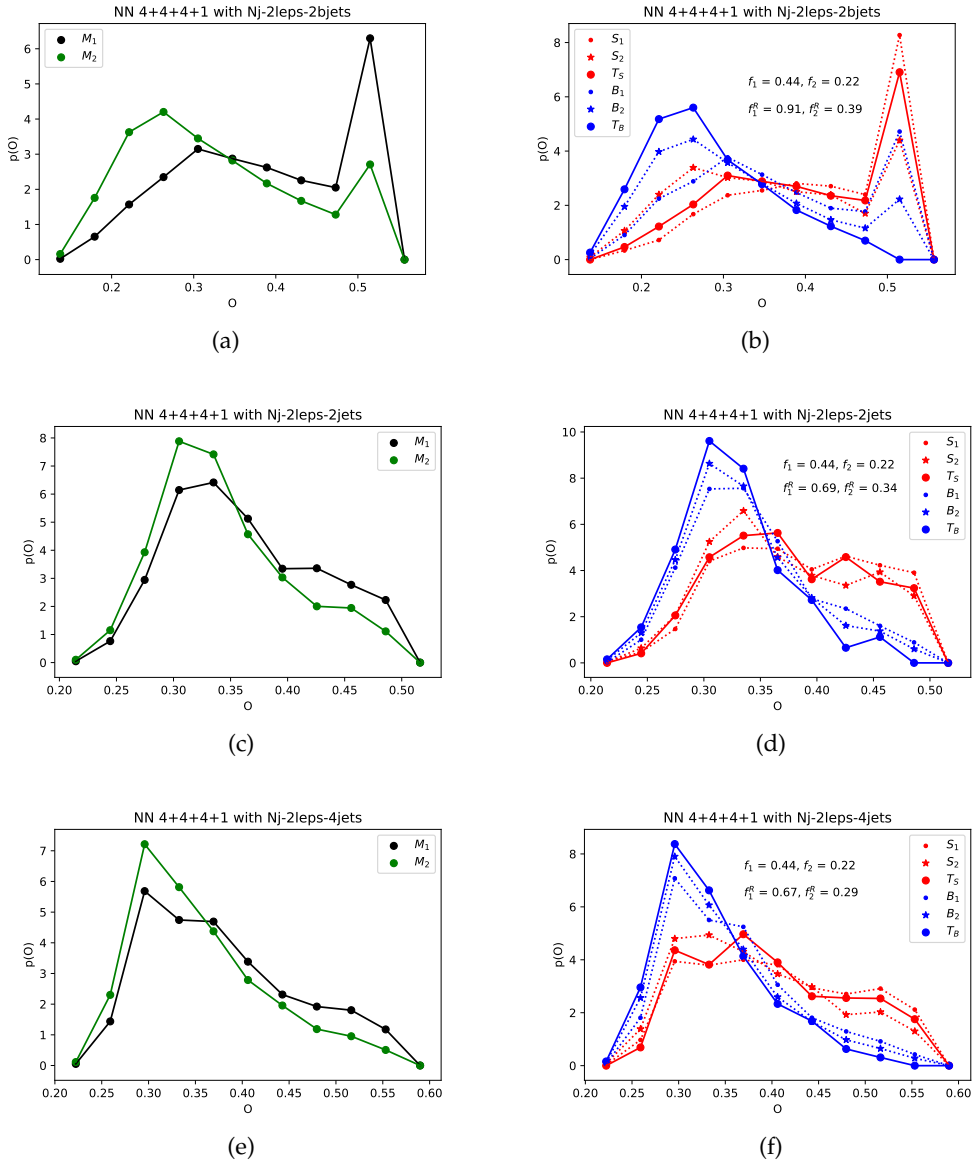



Figure 40: Neural Network outputs with three layers of four neurons preceding a single neuron layer fed with three different set of observables with the label representing whether they belong to $M_1$ or $M_2$. Left panels: Event distribution according to the trained NN output for the events in $M_1$ and $M_2$. Right panel: Reconstructed signal ($S_R$) and background ($B_R$) distributions and underlying truth-level signal and background distributions for $M_1$ and $M_2$ samples ($S_1$, $S_2$, $B_1$ and $B_2$). See discussion in text.

being below $\mathcal{O}(10^{-2})$, makes it challenging to correctly apply the algorithm. Moreover, even in the hypothetical case that new cuts could increase S/B we find that using $N_b$ to define the samples and $N_j$ to demix yields different underlying background distributions, which breaks the H2 hypothesis and therefore spoils the results. In fact, since in this case $t\bar{t}b\bar{b}$ is among the main backgrounds, the relative contribution of this background to the background topic would change considerably between the $N_b = 2$ and $N_b = 3$ samples. Despite these difficulties in approaching this channel with the Demixer algorithm, we still find it an interesting goal, since having a less Monte Carlo

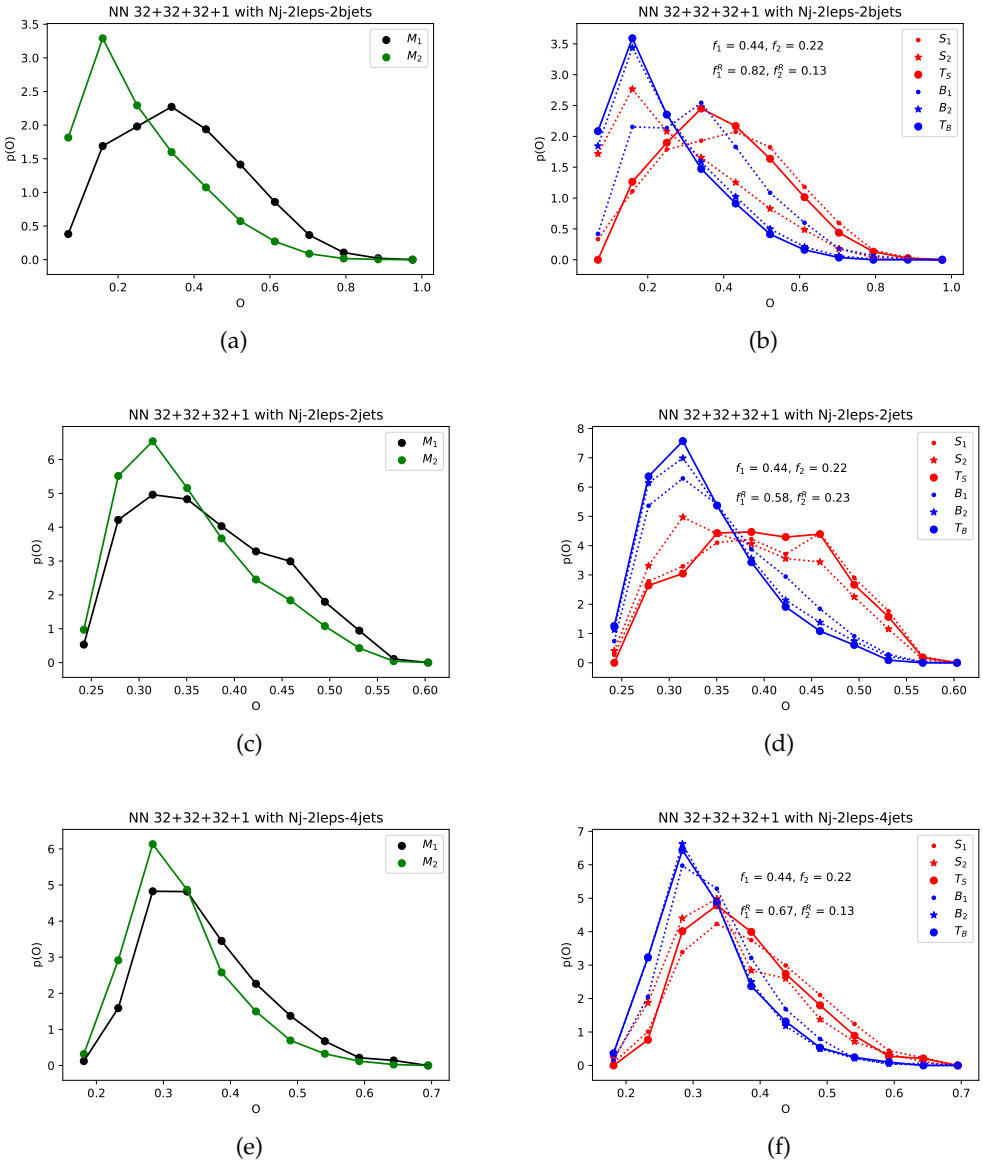



Figure 41: Idem as Fig. 40, but for a more complex 32+32+32+1 NN.

dependent prediction of backgrounds such as $t\bar{t}b\bar{b}$ is an important avenue in 4-top and in heavy flavour physics.

In Subsection 5.2.1 we have used b-tagging information through $N_b$ to define the samples $M_i$, and $N_j$ to demix. An inversion of the roles of these variables would be an interesting study. Of course, in this case one could not demix using $N_b$ since it can take very few discrete values. In fact, to explore an inversion of roles one should define the samples as –for instance– $M_1$: $N_j > 7$ (signal enhanced) and $M_2$: $5 \leqslant N_j \leqslant 7$ (signal suppressed), and define a continuous variable considering the probability of having b-tags within the jets. This variable could be, for instance, the sum of the MV2c20 [206] variable[1] of the four leading jets. This type of variables would

1 This and/or similar variables, and its subsequent development, are the optimal variables on which the probability density of a light-jet and a b-jet are best discriminated. These are the variables used to define the b-tagging working points.

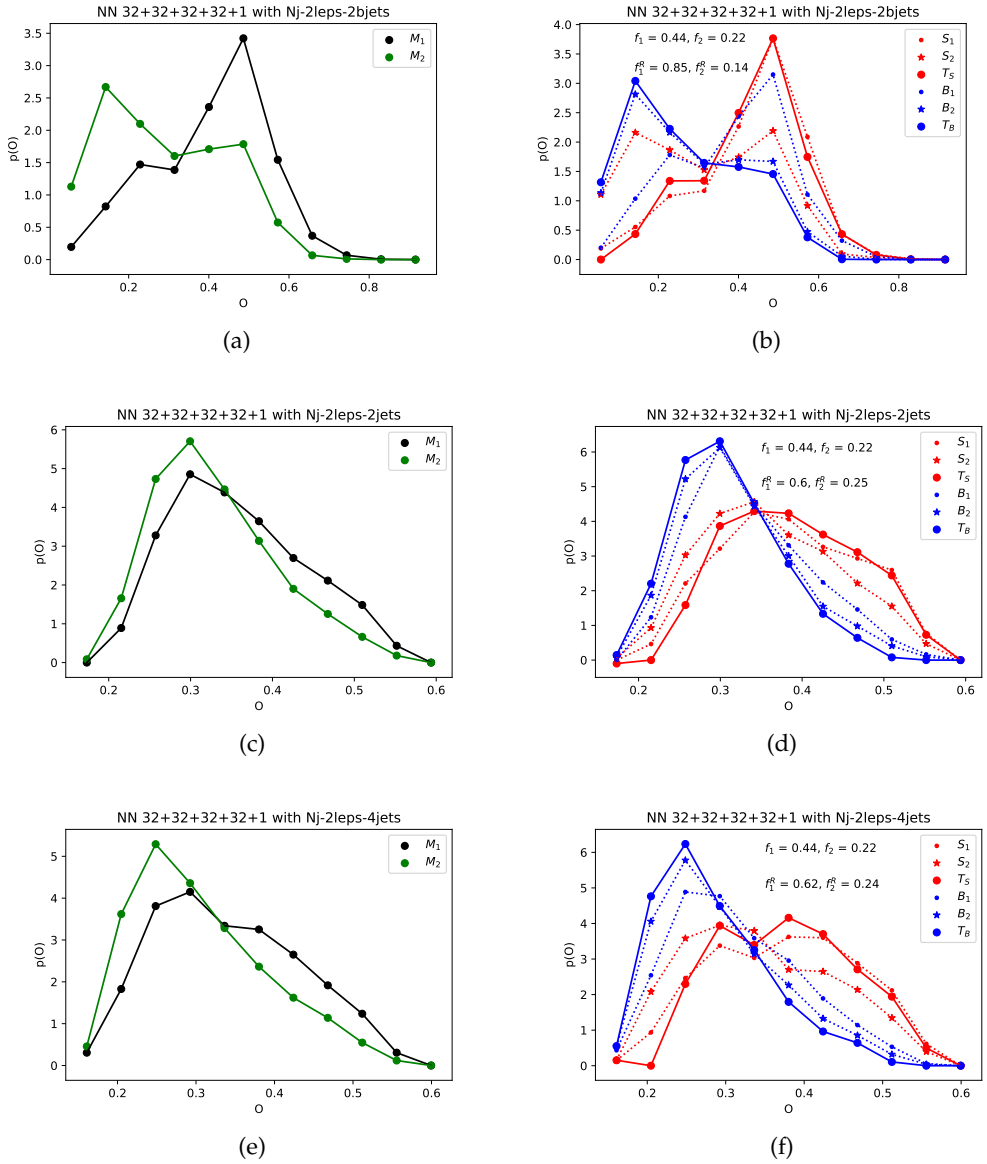



Figure 42: Idem as Fig. 40, but for a more complex and deeper 32+32+32+32+1 NN.

be more suited to the Demixer algorithm and would yield a different performance when extracting the underlying distributions and the signal fractions. In particular, since $N_b$ has a larger discriminating power between signal and background than $N_j$, one could expect a better performance of the Demixer algorithm.

We have also implemented Machine Learning techniques to explore the ability of a Neural Network to best discriminate $M_1$ from $M_2$, and therefore signal from background. The bottleneck in this direction is to obtain a Neural Network whose underlying output distributions remain as similar as possible between $M_1$ and $M_2$, otherwise the breaking of the H2 hypothesis spoils any improvement that could be achieved through deep learning. We have discussed different architectures and choice of parameters in Subsection 5.2.2 where we show the demixing result for each case.

Losing physical interpretation on the demixing variable to enhance the discriminating power yields potential issues in guaranteeing the H2 hypothesis. This in turn is

translated to a more challenging topic reconstruction. When comparing the CWoLA algorithm to the Demixer performance using $N_j$ alone, we find only moderate improvements. This is because $N_j$ is the main discriminator between signal and background at this level. It is also worth noting that the Demixer algorithm is a data-driven technique that does not require training and, using eight i7 CPU cores, it performs with times of order $\mathcal{O}(10^{-3})$ seconds while the CWoLA algorithm requires times of order $\mathcal{O}(10^{2})$ seconds for a simple architecture and a relatively small dataset as implemented in Subsection 5.2.2. One of the reasons for this difference is the small number of bins over which we do the demixing. Nevertheless, we consider that more work in this direction, including a continuous variable for the b-tagging information, could improve the results presented in this Chapter.

We have seen along this Chapter how the algorithm is very sensitive to the H2 hypothesis. Since we are including different backgrounds altogether in the background topic, this makes it easier to break this hypothesis by having different underlying distribution in each sample. This is in particular a crucial failure for the mono-lepton channel briefly discussed above. To tackle this issue, it would be interesting to consider a generalization of the Demixer algorithm that can distinguish between the backgrounds themselves. Work in this direction can be seen in the Factorized Topic Modeling section of Ref. [98].

We end this Section with a brief discussion on how the obtained results using the Demixer algorithm on 4-top could be used to reduce the impact of Monte Carlo simulations and its tuning in the extraction of an absolute physical quantity such as the 4-top production cross section, $\sigma(t\bar{t}t\bar{t})$. As discussed at the beginning of this Chapter, we consider that it is not possible to avoid a Monte Carlo dependence to measure a quantity such as $\sigma(t\bar{t}t\bar{t})$, which can be compared to theoretical matrix level calculations. However, reducing the impact of these predictions and tunings is a crucial task in the 4-top final state which has many complex ingredients such as ISR/FSR, hadronization, jet reconstruction and isolation, among others, which still need to be further understood.

The Demixer algorithm applied to 4-top as implemented in the previous Sections yields a reasonably reliable distribution for the background topic, since the background fairly satisfies having an anchor bin and has more purity in both $M_i$ samples. This is an important result because it provides a Monte Carlo independent prediction for the background in a region where signal is expected. Therefore, one would be implementing a data-driven Monte Carlo tuning for the background in the signal region: tune the Monte Carlo pipeline to reproduce the background shape in $N_j$ extracted from the Demixer algorithm. In contrast to usual Monte Carlo tuning in control regions, where one needs to assume there is no signal, this kind of tuning has the advantage that it does not require an extrapolation to different regions in parameter space while also accounting for the presence of signal in the region. Once this is performed, one could use this tuned Monte Carlo to predict the $\kappa_{SB}$ and from this number extract the shape of the signal distribution in $N_j$ from the Demixer algorithm. A comparison of the signal shape predicted by the tuned Monte Carlo and the shape predicted by the Demixer once the $\kappa_{SB}$ has been determined, would be a measure of the success of the method. In such a case, one could extract the fractions of signal and background in $M_i$ and rely on this tuned Monte Carlo to extract the $t\bar{t}t\bar{t}$ production cross section.

## 5.4 CONCLUSIONS

We have studied the Demixer algorithm as well as some of its limitations, and applied it to the 4-top signal and its background at the LHC, where many of the required hypotheses are not fully satisfied.

In real case scenarios the requirements for applying the Demixer algorithm are usually not exactly satisfied. We have analyzed how the outcome of this algorithm is affected when some of the required assumptions break down. We find that, to some extent, the method is still useful. We have shown explicitly how the topics reconstruction and fraction estimations may shift because of not having equal underlying distributions in the initial mixture samples and/or there is not mutual irreducibility. The study is not exhaustive, and we conclude that more work in this direction is needed to better understand the Demixer algorithm and its scope beyond ideal cases.

We have implemented the Demixer algorithm for the 4-top final state and its main backgrounds. We have used the same sign dilepton channel which assures a reasonable S/B and has the special feature that all backgrounds have no more than two b-jets in their final state. The implementation has been made using simulated events up to detector level. We have defined the two mixture samples using the number of b-tags, $N_b = 2$ and $N_b = 3$, while we have used $N_j$ to demix the samples. We have also essayed to demix using Machine Learning through a Neural Network which best discriminates the mixed samples, as in CWoLA. We find that Neural Network results can be slightly better, but at the price of considerably hiding the clarity of just using $N_j$. Neglecting statistic and systematic uncertainties we have shown that the reconstruction of signal and background topics is as predicted by the Demixer algorithm beyond the ideal case: *i)* A lack of anchor bin in signal causes that only the background is correctly reconstructed; *ii)* The signal is correctly reconstructed provided the corresponding reducibility $\kappa$-factor, which requires input beyond data-driven; and *iii)* The estimation of the fractions and reconstruction of the topics yield values somewhat shifted from its true values due to slight differences in the underlying distributions. The Demixer algorithm in the described framework without using Monte Carlo inputs predicts the fraction of signal in the mixture samples with a misestimation of approximately $\sim 30\%$.

We have discussed some possible future directions and improvements regarding the results of this Chapter. For example, one could invert the roles of $N_b$ and $N_j$ when defining the samples and demixing, replacing $N_b$ by a more continuous b-tagging variable. We have proposed to consider the Demixer algorithm to obtain a Monte Carlo independent prediction for the 4-top background, which would allow to tune the Monte Carlo parameters for the background prediction in the signal region. This tuned Monte Carlo could be used as a new tool to extract physical quantities from the signal region, as for instance the 4-top production cross section, achieving the purposed goal of reducing the impact of Monte Carlo in the extraction of physical quantities.

4-top is a very challenging final state within any framework and the Demixer algorithm is not the exception. The actual implementation of the algorithm for 4-top at the LHC along the lines presented in this Chapter would require the full LHC Run 3 dataset, and still many experimental aspects and uncertainties would have to be further analyzed. Among the main issues to be addressed in a real implementation, we can mention that a better discriminating variable than $N_j$ that could provide an improved differentiation between the mixture samples would be a crucial milestone. We

present the results in this Chapter as an alternative step towards reducing the impact of Monte Carlo generators in the extraction of physical quantities in 4-top physics.

# 6

# LDA FOR LHC

In this Chapter we build upon a technique proposed in Ref. [100] to classify jets and events within a single mixed event sample. The tools used by this technique have been developed in a branch of Machine Learning called generative statistical modeling [71] with the primary goal of identifying emergent themes in collections of documents. To accomplish these, the generative models infer the hidden (or latent) structure of a document corpus using posterior Bayesian inference based on word and theme co-occurences [207–213]. Using the example of jet substructure observables based on the clustering history, Ref. [100] provides an statistical mixed membership model of jet substructure. In particular, using the model of Latent Dirichlet Allocation (LDA) [208], which can be solved efficiently using Variational Inference (VI) [214] techniques, Ref. [100] was able to define robust parametric jet and event classifiers.

In the present Chapter we provide further details on this approach, building upon the basic assumptions and premises about the relevant measurements/observables and their statistical modeling in order to construct relevant and practical generative Bayesian models for particle physics event classification, in particular LDA. We also provide further justification for why jet clustering history observables are a particularly interesting and applicable example for these methods. Finally, we perform a systematic study of the parametric and Bayesian prior dependence and performance of VI and LDA, respectively, based on two representative examples of boosted $t\bar{t}$ events and events containing hypothetical Beyond the Standard Model (BSM) boosted colour neutral, but hadronically decaying scalars [74, 81, 87, 215, 216]. In particular, we identify *perplexity* as a robust measure of LDA and VI performance which does not rely on access to labelled data while being correlated with traditional classification performance measures like tagging efficiency and mistag rate. We then use this unsupervised criteria to identify the most suitable parameter and prior ranges for the example datasets.

This Chapter is based on work done in collaboration with Barry M. Dillon, Darius A. Faroughy and Jernej F. Kamenik in Ref. [217] and is structured as follows. In Section 6.1 we outline the statistical premises and introduce the Bayesian generative models upon which LDA is based. We also provide details of LDA training and inference methods and how they can be applied to event classification. In Section 6.2 we apply the general framework to the multi-jet event data in the form of jet clustering history observables and discuss the most appropriate data representations. The benchmark event samples used for our study are introduced in Section 6.3, where we also discuss the data preparation steps that need to be considered when using LDA. Section 6.4 contains the main results of our systematic study of LDA based classification methods applied to the example datasets. Finally we summarize our conclusions and provide an outlook in Section 6.6. All necessary code to reproduce our results is available at Ref. [218].

## 6.1 PROBABILISTIC GENERATIVE MODELLING FOR COLLIDER EXPERIMENTS

The goal in high energy collider experiments is to gain understanding of the hard collisions resulting from the underlying physical processes taking place at very high energies during the events. Each event can result in anywhere from $\mathcal{O}(10)$ to $\mathcal{O}(1000)$ particles being detected away from the beamline and the detector records the energy, momentum, and tracking information of these particles. One must then analyse this high-dimensional dataset and compare it to what is expected from theoretical predictions in order to gain an understanding of the underlying physics.

To perform this analysis in practice the dimensionality of the dataset must be drastically reduced. The choice of dimensionality reduction depends on the type of underlying physics one wants to study and typically involves some combination of jet clustering and grooming, pile-up subtraction, the use of certain high-level observables, and performing cuts to remove unnecessary elements of the dataset. Once the dataset has been processed and the relevant high-level observables have been obtained, a statistical analysis comparing the measurements with the theoretical predictions can quantify how much a particular physics model agrees with the data.

Bayesian probabilistic generative modeling is an unsupervised machine learning approach in which one constructs a probabilistic model for a dataset, and then uses approximate inference techniques to estimate the parameters of this probabilistic model directly from the data. If the probabilistic model is a good approximation to how the data was actually generated, this in turn allows to identify patterns in the dataset. In our case these patterns could contain important information on the underlying physical processes registered in collider events.

In order to study our dataset, it is useful to think of a single event $e_j$ $(j = 1, 2, \ldots, N_e)$ as the finite list of measurements $e_j = \{o_{j,1}, o_{j,2}, \ldots\}$, where $o_{j,i}$ $(i = 1, 2, \ldots, N_j)$ are in general functions (or mappings) of the relevant multi-particle phase-space, that compose it. This choice of data representation allows us to construct a model for the events by supposing that the measurements have been sampled from a (presumably complicated) joint probability distribution $p(e_j) = p(o_{j,1}, o_{j,2}, \ldots)$. This is the starting point for the unsupervised analysis techniques used in this Chapter.

Writing a general statistical model describing the generative process of events is usually not possible in practice. To proceed, it is necessary to impose a set of simplifying assumptions on the joint probability distribution. The functional dependence of this distribution on $o_{j,i}$ must be flexible enough in order to account for the multiple physical processes manifest in each event, but it also must be simple enough such that efficient inference techniques can be implemented. In order to model the events using the techniques described in this Chapter, the phase-space observables furthermore need to be labeled and binned so that the possible measurements $o_{j,i}$ are discrete and finite in number. This requirement allows us to construct probabilistic models based on multinomial distributions that describe the occurrence of the measurement bins $o_{j,i}$ in events. If we were to consider unbinned observables then $p(e_j)$ would be constructed from continuous probability distributions. Moreover, given that the measurements we work with are coarse-grained due to detector granularity and reconstruction uncertainties, there is a physical justification to the binning assumption.

We give a brief introduction to probabilistic generative models in general in Subsection 6.1.1 and to the specific case of LDA in Subsection 6.1.2.

### 6.1.1 *Probabilistic generative models*

In order to introduce the concepts and models used in this Chapter we discuss two different models for $p(e_j)$ of increasing complexity: mixture models, and mixed-membership models. The starting point for the construction of these models is de Finetti's representation theorem [219], which states that if the measurements in the joint probability distribution are exchangeable then the measurements are conditionally independent given some latent variables. The exchangeability requirement simply means that the joint probability distribution should be invariant under a re-ordering or exchanging of the measurements. Note that measurement exchangeability is not to be confused with measurement independence. The former is a weaker condition that leads to more flexible probabilistic models capable of capturing complicated hierarchical patterns in the data. Formally, the theorem states that the joint probability distribution can be written as

$$p(o_{j,1}, o_{j,2}, ..., o_{j,N_j}) = \int_\Theta d\theta\, p(\theta) \prod_{i=1}^{N_j} p(o_{j,i}|\theta)\,, \tag{88}$$

where $\theta$ ($\Theta$) represents a hidden latent parameter (space) which is marginalised over in the joint probability. From the $p(o_{j,i}|\theta)$ factor on the right-hand side of the above equation we see that the conditional independence of the different measurements is manifest at the cost of the explicit conditional dependence on the latent space variables. When viewing this result from a Bayesian perspective, the probability functions $p(o|\theta)$ represent the likelihood functions while the function $p(\theta)$ outside of the product acts as a prior distribution over the latent space. This theorem underpins probabilistic models such as mixture models and mixed-membership models, which we discuss in the following paragraphs.

*Mixture models*

We present one of the simplest probabilistic models for a sample of collider events: the mixture model. This model assumes there are $T$ different probability distributions over the measurement bins, represented by $p(o|t, \beta)$ for $t = 1, \ldots, T$, that can generate a given event $e_j$. Because of the discrete neature of our observable-space, these probability distributions are $M$-dimensional multinomials with $M$ the number of bins in said observable-space. The parameters of these multinomials are represented by the elements of a $T \times M$ dimensional matrix, $\beta_{t,m}$ having the property that $\sum_{m=1}^{M} \beta_{t,m} = 1$ for all $t$. Each row in $\beta_{t,m}$ contains the parameters of the multinomial associated to one of the $T$ probability distributions.

A key feature of mixture models is that they assume measurements in a single event have been sampled from just one of these $T$ multinomial distributions. So for each event one of the $T$ distributions is selected from another multinomial probability distribution $p(t|\omega)$, where $\omega = (\omega_1, \ldots, \omega_T)$ are the probabilities to select each $T$. They satisfy $0 \leqslant \omega_t \leqslant 1$ and $\sum_t^T \omega_t = 1$. In this Chapter we refer to each $T$ latent multinomial distribution as a 'theme', a term inherited from the field of topic modeling where these methods were popularised. The $\omega$ parameters are thus referred to as theme weights.

It is useful to describe the probabilistic model in terms of a generative process, outlining the underlying assumption on how the events were generated. The generative process for a collection of events in a mixture model goes as follows:

1. Randomly sample a theme $t_j \sim p(t|\omega)$.
2. Randomly sample a measurement $o_{j,i} \sim p(o|t_j, \beta)$.
3. Repeat step (ii) for each measurement in the event.
4. Repeat steps (i)-(iii) for each event in the sample.

The mathematical structure of the model can be realised by taking the representation of the joint probability distribution written as in Eq. 88 and defining $\theta \in \mathbb{R}$ with a prior distribution over the latent space $p(\theta) = \sum_t^T p(\theta|\omega)\delta(\theta - t)$ where $p(\theta|\omega)$ is a density such that $p(\theta = t|\omega) = \omega_t$. This leads us to the form

$$p(e_j) = \sum_{t_j=1}^{T} p(t_j|\omega) \prod_{i=1}^{N_j} p(o_{j,i}|t_j, \beta)\,. \tag{89}$$

The kind of generative processes described here can be thought of as graphical models [71], which allows us to represent them with a plate diagram: the unobserved variables (latent random variables and model parameters) are represented by white circles, observed data (measurements) are represented by shaded circles, while the conditional dependencies and independent identically distributed i.i.d samplings are represented by arrows. To indicate that certain steps in the generative process are repeated, a labelled box or plate is drawn around the relevant parts of the diagram, with the integer label representing the number of times these steps are to be repeated.

The graphical mixture model corresponding to the generative process described in Eq. 89 is shown in the upper diagram in Fig. 43. We see there that the free parameters of the mixture model, the theme proportions $\omega_t$ and the multinomial probabilities $\beta_{t,m}$, are located outside all plates and thus have to be defined for the whole event sample in order to initiate the generative process that leads to the measurements $o_{j,i}$ in the inner-most plate.

In collider physics, event samples arise from a statistical mixture of multiple underlying physical scattering process and each event is a result of one such particular scattering process. Once the appropiate differential cross-sections are binned, the scattering processes can be identified with themes in a multinomial mixture model as described above. Traditionally, the weights $\omega$ are computed from first principles using a combination of Quantum Field Theory, Monte Carlo event generators tuned to data and experimental knowledge of the detector response. The work detailed in Chapter 5 and Refs. [72, 82] are examples of semi-supervised classification with mixture models, classifying event samples where the mixture proportions of the themes are a priori unknown.

There are several drawbacks when using mixture models for (unsupervised) event classification tasks. These come from the assumption that all measurements in an individual event are drawn from one theme. The main (related) issues are the following:

- Measurements on a single collider event typically receive contributions from many sources, for example in measuring $t\bar{t}$ production it is inevitable that much of the measurements will be of soft QCD radiation rather than the hard decay products of the top quarks. Mixture models fail by design to differentiate between different underlying processes in individual events.
- Mixture models are not well suited for modeling datasets where events generated from different themes share common features. This is not because of the

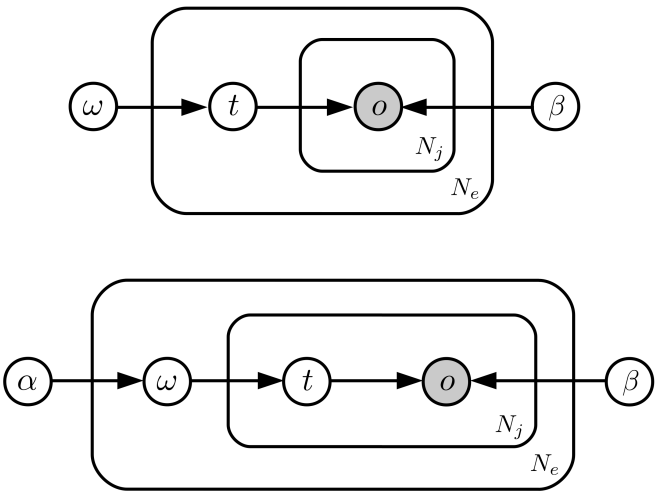



Figure 43: The graphical models representing the generative process for $N_j$ measurements in $N_e$ events for the mixture model Eq. 89 (upper diagram) and the mixed-membership model Eq. 91 (lower diagram). See text for details.

mixture models themselves but because of approximate inference techniques required to extract the themes, see Ref. [208].

In general, mixtures are useful representations of the data if the mixing proportions $\omega$ can be computed from first principles or estimated with other means (such as in (semi)supervised ML), but tend to be less suitable if the weights are unknown, as is the case with unsupervised searches. Mixed-membership models, which address these issues in an efficient way, are introduced next.

*Mixed-membership models*

Mixed-membership models assume $T$ themes as in the mixture model but consider that a single event is generated from a mixture of themes rather than from a single theme. Each event $e_j$ now has its own theme weights $\omega_j = (\omega_{j,1}, \ldots, \omega_{j,T})$. Instead of parameters, these theme weights are now latent variables of the model. They are sampled from a prior distribution $p(\omega|\alpha)$, with $\alpha$ being the parameters of the distribution. This prior is in general defined over the $(T-1)$–dimensional simplex describing the space of all theme weights (which is the space of $T$-vectors with positive entries that sum up to one). The generative process for a mixed-membership model goes as follows:

1. Randomly sample a set of $T$ theme proportions from the prior, $\omega_j \sim p(\omega|\alpha)$.
2. Randomly sample a theme $t_i \sim p(t|\omega_j)$.
3. Randomly sample a measurement $o_{j,i} \sim p(o|t_i, \beta)$.
4. Repeat steps (ii)-(iii) for each measurement in the event.
5. Repeat steps (i)-(iv) for each event in the sample.

We can derive the mixed-membership representation of the joint probability by taking Eq. 88 and assigning

$$p(o_{j,i}|\omega_j) = \sum_{t_i=1}^{T} p(t_i|\omega_j)\; p(o_{j,i}|t_i, \beta), \tag{90}$$

where $p(t_i|\omega_j) = \omega_{j,t_i}$. Therefore the mixed-membership model probability for an event is defined by

$$p(e_j) = \int_\Omega d\omega_j\; p(\omega_j|\alpha) \prod_{i=1}^{N_j} \sum_{t_i=1}^{T} p(t_i|\omega_j)\; p(o_{j,i}|t_i, \beta)\,, \tag{91}$$

where $\Omega$ is the $(T-1)$–dimensional simplex. Note the slight change in notation: the latent space variable $\theta$ from Eq. 88 has been replaced with $\theta \to \omega_j$ (and $\Theta \to \Omega$) to keep the notation for mixed-membership models in line with the notation for mixture models. The generative process for the mixed-membership model is shown in the lower diagram of Fig. 43. In comparison to the mixture model plate diagram, notice that the theme selection step is now inside the event plate indicating the mixed-membership nature of the model. The free parameters of the mixed-membership model are $\alpha$ from the prior and the multinomial probabilities $\beta_{t,m}$ of the themes.

The mixed-membership assumption allows us to model events sharing similar features, which was a challenge when using mixture models. It is therefore possible to model events that are much more heterogeneous. It is also clear that the model can now describe events where measurements receive contributions from multiple sources, accommodated now by each event containing measurements sampled from different themes. Going back to collider physics, these themes represent the different archetypes underlying the different physical processes. For example, $t\bar{t}$ production has measurements sampled from the hard $t\bar{t}$ archetype and from the soft QCD archetype.

### 6.1.2 *Latent Dirichlet Allocation*

LDA is an example of a mixed-membership model, being defined by a clever choice of the priors over $w$ ($p(\omega|\alpha)$) and, when smoothed, $\beta$ ($p(\beta|\eta)$). This choice is very useful when perform inginference of the model parameters over a corpus to obtain the underlying themes and weights. Parameter inference is discussed in detail in Subsection 6.1.3. In particular, choosing the prior $p(\omega|\alpha)$ to be the conjugate distribution to the likelihood function makes the parameter inference easier. For mixed-membership models with a multinomial likelihood, as we have here, the conjugate prior is the Dirichlet distribution $\mathcal{D}(\cdot|\alpha)$. Choosing $p(\omega|\alpha)$ in Eq. 91 to be a Dirichlet distribution leads us to LDA [208, 210].

The Dirichlet distribution defined over the simplex $\Omega$ is a multivariate generalisation of the beta distribution over the unit interval $[0, 1]$, reducing to the Beta distribution for $T = 2$. It is in fact a parametric family of distributions defined by $T$ positive non-zero parameters $\alpha = \alpha_1, \ldots, \alpha_T$ and has the explicit form

$$\mathcal{D}(\omega|\alpha) = \frac{\Gamma\left(\sum_{t=1}^{T} \alpha_t\right)}{\prod_{t=1}^{T} \Gamma(\alpha_t)} \prod_{t=1}^{T} \omega_t^{\alpha_t - 1}\,, \tag{92}$$

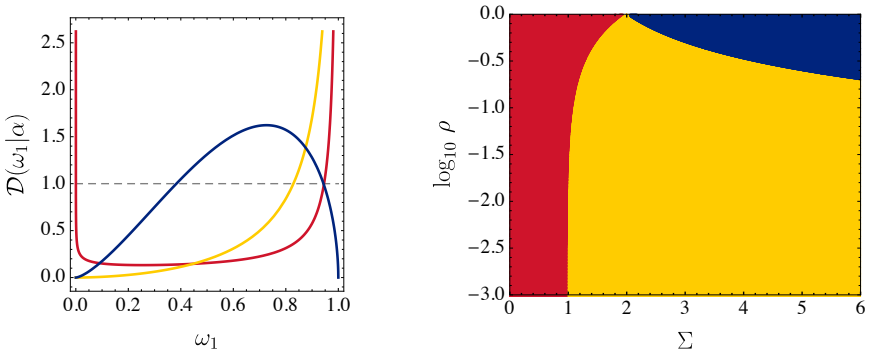



Figure 44: (Left) Three representative Dirichlet priors for $T = 2$ over the unit interval drawn in different coloured full lines. The uniform prior $\mathcal{D}(\omega_1|1,1)$ is also shown in dashed grey. (Right) Regions with the same prior shapes projected onto the $(\Sigma, \rho)$-plane defined in Eq. 94. The distinct shape inside each coloured region is represented by one distribution in the left panel with matching colour codes.

where $\Gamma(x)$ is the gamma function. In LDA, the Dirichlet prior encodes prior information on how we expect the themes to contribute both to individual events and to the whole sample of events. It does this by influencing the possible proportions $\omega_j$ selected in the generative process. For example a particular choice of parameters $(\alpha)$ could define a model in which one particular theme contributes much less to individual events than another, or it could define a model in which some events are composed almost exclusively of one theme while others are more equal mixtures of several themes.

As mentioned in the introduction to this Chapter, we are concerned solely with scenarios in which only two themes are relevant. We therefore focus on the $T=2$ case where the Dirichlet prior reduces to a Beta distribution. For each event we sample a variable $\omega_1$ from the Dirichlet (Beta) distribution representing the proportion of the first theme $p(o|1,\beta)$, while the proportion of the second theme $p(o|2,\beta)$ is given by $\omega_2 = 1-\omega_1$. In this two-theme scenario, the analytical form of the Dirichlet is given by

$$\mathcal{D}(\omega_1|\alpha_1,\alpha_2) = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\,\Gamma(\alpha_2)}\,\omega_1^{\alpha_1-1}(1-\omega_1)^{\alpha_2-1}, \tag{93}$$

where for now we drop the j subscript labelling the event. When inspecting the above distribution family for different values of the $\alpha$ parameters, one identifies several cases that give rise to different types of distribution shapes. These different shapes encode different assumptions about the event data. For instance $\mathcal{D}(\omega_1|1,1)$ corresponds to the flat distribution over the unit interval and would describe events for which the occurrence of either theme in an event is completely random (shown in gray dashed line in Fig. 44). The other, more interesting, shapes are the following:

1. $\alpha_1 < 1,\ \alpha_2 < 1$: bi-modal distributions (shaded in red in Fig. 44) with two maximums at the boundaries of the unit interval ($\omega_1 = 0$ and $\omega_1 = 1$). Physically, this scenario describes samples for which one group of events have measurements predominantly sampled from the first theme, and another group for which measurements are mostly sampled from the second theme. The relative size between each group of events is controlled by the ratio $\alpha_2/\alpha_1$.

2. $\alpha_1 > 1,\ \alpha_2 < 1$: uni-modal distributions with a maximum located at one boundary of the interval and the distribution tail stretching towards the opposite boundary (shaded in yellow in Fig. 44). In this case we expect most events to be generated mostly by one predominant theme.

3. $\alpha_1 > 1,\ \alpha_2 > 1$: uni-modal distributions with one maximum located at $\omega_1 = \frac{\alpha_1 - 1}{\alpha_1 + \alpha_2 - 2}$ and two tails stretching towards both boundaries of the interval (shaded in blue in Fig. 44). In this case we expect the bulk of events to be generated by non-negligible mixtures of both themes, with very few events where just one theme completely dominates. However the exact distribution depends strongly on the hierarchy between $\alpha_1$ and $\alpha_2$.

In the following sections we will rely on a useful re-parametrization of the Dirichlet where we trade the $(\alpha_1, \alpha_2)$ parameters for $(\Sigma, \rho)$ defined as

$$\Sigma \equiv \alpha_1 + \alpha_2\,, \quad \rho \equiv \frac{\alpha_2}{\alpha_1}\,. \tag{94}$$

By convention we have fixed here $\alpha_2 \leqslant \alpha_1$ and thus $0 < \rho \leqslant 1$. The $\rho$ parameter controls the asymmetry in the shape of the Dirichlet distribution. In Fig. 44 (right) we present a visualisation of the different shapes taken by the Dirichlet distribution, in terms of these $\rho$ and $\Sigma$ parameters. The smaller $\rho$ is, events are more likely to be composed of measurements drawn from the first theme ($t = 1$). A way to see this is by considering the expectation for sampling the themes from the Dirichlet during one measurement sampling:

$$\begin{aligned} \mathbb{E}_{\mathcal{D}}\left[p_1\omega_1 + p_2(1-\omega_1)\right] &= \int_0^1 d\omega_1\ \mathcal{D}(\omega_1|\alpha)\left[\omega_1 p_1 + (1-\omega_1)p_2\right] \\ &= \frac{p_1 + \rho\, p_2}{1+\rho}\,, \end{aligned} \tag{95}$$

where $p_t$ are shorthand for the theme multinomials $p(o|t, \beta_{t,m})$ and $\mathbb{E}_{\mathcal{D}}[\cdot]$ denotes the expectation with respect to the Dirichlet distribution. To derive this we have used the expression for the mean value

$$\mu = \mathbb{E}_{\mathcal{D}}[\omega_1] = \frac{1}{1+\rho}.$$

This indicates that in the limit $\rho \to 0$ of asymmetric Dirichlet priors there will be a prevalence of the first theme over the second theme when sampling measurements for an event, while in the limit $\rho \to 1$ the priors become symmetric and events will tend on average to have measurements coming from both themes in similar proportions.

The parameter $\Sigma$, on the other hand, controls to what degree individual events in the model are described by mixtures of themes for a fixed value of the asymmetry parameter $\rho$. In other words, $\Sigma$ controls to what degree the model is a mixed-membership rather than just a mixture model. For large $\Sigma$ we expect that events are generated from mixtures of both themes, whereas for $\Sigma \ll 1$ we expect that events are generated from pre-dominantly one theme.

In fact, it is known that the Beta distribution will approach the Bernoulli distribution in the limit of $\Sigma \to 0$ with fixed $\rho$. In general, a Dirichlet for $T$ themes will approach a T-dimensional multinomial distribution in the limit $\sum_{t=1}^{T} \alpha_t \to 0$ [220]. In this limit the Bernoulli probability $p$ is equal to the expectation value of the Dirichlet, $\frac{1}{1+\rho}$. Therefore in the $\Sigma \ll 1$ limit each event is approximately generated by just one theme,

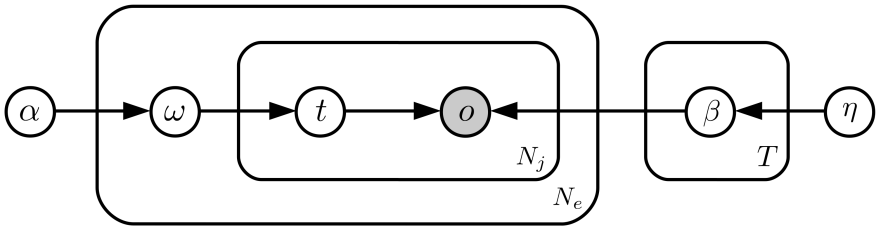



Figure 45: The graphical model of smoothed LDA.

and the LDA mixed-membership model tends to the mixture model described above. In this limit, when you sample $(\omega_{j,1}, \omega_{j,2})$ from the Dirichlet for every event, the only weights that have non-zero probability are $(\omega_{j,1} = 1, \omega_{j,2} = 0)$ and $(\omega_{j,1} = 0, \omega_{j,2} = 1)$, where the probabilities for selecting each of these from the Dirichlet is $1/(1+\rho)$ and $\rho/(1+\rho)$, respectively. In this mixture model limit $\rho$ takes on the role of the ratio of theme weights $\omega_1/\omega_2$. In Fig. 44 we can thus identify the boundary at the y-axis as a mixture model with $\omega_1/\omega_2 = \rho$.

On the other hand, insight into the different weights for each theme across the sample can be gained by observing that the variance of the theme sampling during one measurement sampling is

$$\begin{aligned} \mathrm{Var}_{\mathcal{D}}\left[p_1\omega_1 + p_2(1-\omega_1)\right] &= \int_0^1 d\omega_1 \; \mathcal{D}(\omega_1|\alpha)\left[\omega_1 p_1 + (1-\omega_1)p_2 - \frac{p_1+\rho p_2}{1+\rho}\right]^2 \\ &= \frac{\rho}{(1+\rho)^2}\frac{(p_1-p_2)^2}{\Sigma(\Sigma+1)}, \end{aligned} \tag{96}$$

which indicates that in the limit of large $\Sigma$, for fixed positive $\rho$ and different measurement probabilities, the weights for each theme are very similar across the sample.

The event samples we analyse can contain anywhere from $\mathcal{O}(10^3)$ to $\mathcal{O}(10^6)$ events and the number of unique measurements can also be very large. This means that the data will be very sparse in some regions of the observable-space, with many $o_{j,i}$ appearing rarely in the sample. This can lead to issues in the inference procedure, with these rare measurements being assigned zero probability in the themes, which then leads to problems during the classification of events. This issue can be solved by the so-called 'smoothing' of the themes [208]. Smoothing involves placing a M-dimensional Dirichlet prior on the variables of the theme probability distributions, ensuring that no measurement can have a zero probability. The generative process is then augmented as shown in the plate diagram in Fig. 45. In this Chapter we consider smoothed LDA and we fix each of the $M-1$ parameters of each theme Dirichlet prior to $1/M$ as default. In the following, whenever we mention LDA we are referring to this smoothed LDA with default theme priors.

### 6.1.3 *Variational inference*

Ultimately, the goal is to estimate the posterior distributions for the variables in the LDA model given the observation of experimental data. We can write the joint probability over all events $e = (e_1, \ldots, e_{N_e})$, for the LDA model as

$$p(e,\beta,\omega,t|\alpha,\eta) \propto \prod_{j=1}^{N_e}\prod_{i=1}^{N_j} p(o_{j,i}|t_{j,i},\beta)p(t_{j,i}|\omega_j)p(\omega_j|\alpha)\prod_{t=1}^{T} p(\beta|\eta), \tag{97}$$

where we have not marginalised over the model variables. On the left hand side of this equation, $\omega$ represents the list of theme weights for all events in the sample and $t$ represents the list of topic assignments for each $o_{j,i}$ in all events in the sample. The joint probability is the probability of having generated these events given the LDA model with the themes each being sampled from a Dirichlet parametrized by $\eta$, and the theme weights being sampled per event from a Dirichlet parametrized by $\alpha$. From this we want to approximate the posterior distribution $p(\beta,\omega,t|e)$. Bayes theorem states that this posterior should have the form

$$p(\beta,\omega,t|e) \propto p(e,\beta,\omega,t)/p(e)$$

Although the term in the numerator is calculable, the normalization term known as the evidence presents more difficulties. This evidence is an intractable integral and prevents us from straightforwardly obtaining a closed form expression for the posterior distribution. However, we can obtain an approximation to the true posterior distribution using approximate inference techniques. Following Refs. [214, 221], we implement the Variational Inference (VI) method to approximate the posterior. VI introduces an approximation of the posterior $q(\beta,\omega,t)$

$$p(\beta,\omega,t|e) \simeq q(\beta,\omega,t) \equiv q(\beta)q(\omega)q(t)\,, \tag{98}$$

where $q(\beta,\omega,t)$ is assumed to factorise in each variable, reflecting how these are grouped in LDA. The goal of VI is to obtain the $q(\beta,\omega,t)$ function that best approximates the posterior $p(\xi|e)$. It does so by expressing the logarithm of the evidence as

$$\begin{aligned}\log p(e) &= \int d\xi\; q(\xi)\log\frac{p(e,\xi)}{q(\xi)} + \int d\xi\; q(\xi)\log\frac{q(\xi)}{p(\xi|e)}\\ &= \mathcal{L} + \mathrm{KL}(q(\xi)\big|\big|p(\xi|e))\,,\end{aligned} \tag{99}$$

where we use $\xi$ as a shorthand for the model variables $(\beta,\omega,t)$ and log is the natural logarithm. On the right-hand-side of Eq. 99 we have two terms: the Evidence-Lower-BOund (ELBO) $\mathcal{L}$ and the Kullback-Leibler or KL divergence [222], a positive quantity that reflects the relative entropy between the true posterior and its approximation. The KL divergence vanishes only when $q(\beta,\theta,t) = p(\beta,\theta,t|e)$ and is strictly positive otherwise. The term $\mathcal{L}$ is thus a lower bound on the evidence, which is why it is called ELBO. Although we cannot compute the KL divergence because we cannot compute the posterior, we can compute the joint likelihood and thus the ELBO term. Because the evidence term on the left-hand-side is completely independent of $q(\beta,\omega,t)$, maximising the ELBO is equivalent to minimising the KL divergence between $q(\beta,\omega,t)$ and the posterior. The goal is then to maximise the ELBO with respect to $q(\beta,\omega,t)$ thus finding a $q(\beta,\omega,t)$ which is a good approximation to the posterior. VI gives us a prescription for doing this in mixed-membership models like LDA. As detailed in Ref. [71], for a factorized approximation $q(\xi) = \prod_{i=1} q_i(\xi_i)$ where $\xi_i$ are non-overlapping sets of parameters we can obtain the expression for the optimal solutions $q_i^*(\xi_i)$:

$$\log(q_i^*(\xi_i) = \mathbb{E}_{q^*_{j\neq i}}[\log(p(e,\xi))] + C \tag{100}$$

where the expectation value is taken on all other sets $\xi_j$ following the $q_j^*$ distributions. We thus have a set of equations that represent consistency conditions to be resolved iteratively, much like the Expectation-Maximization algorithm used for Maximum-Likelihood estimates of mixture models.

The LDA model belongs to the conjugate exponential family of models. For these, one can see from Eq. 100 that the terms in the posterior approximation must have the following form:

$$\begin{aligned} q(t_{j,i}) &= \text{Multinomial}(\phi_{j,i}), \quad (j = 1, \ldots, N_e), \ (i = 1, \ldots, N_j), \\ q(\omega_j) &= \text{Dirichlet}(\gamma_{j,t}), \quad (j = 1, \ldots, N_e), \ (t = 1, \ldots, T), \\ q(\beta_t) &= \text{Dirichlet}(\lambda_{t,m}), \quad (t = 1, \ldots, T), \ (m = 1, \ldots, M). \end{aligned} \tag{101}$$

So to optimise $q(\beta, \omega, t)$ we need to maximise the ELBO with respect to the parameters $\phi_{j,i}, \gamma_{j,t}, \Lambda_{t,m}$. Note that there are $T(n_e+M)+N_J$ parameters here, where $N_J$ is the total number of measurements in all events in the sample. Due to the conditional dependencies and the use of conjugate priors specific to LDA, closed form expressions of the parameters that maximise the ELBO through Eq. 100 can be written in terms of each other, as we show below. The VI algorithm then dictates how to update the parameters iteratively such that it converges to a maximum of the ELBO function.

Due to the large number of events from which we infer the parameters of the approximate posterior, the basic VI algorithm is inefficient. To implement this in a way which scales well to large datasets we employ an extension of this algorithm called Stochastic Variational Inference (SVI). This technique uses results from stochastic optimisation methods to speed up the inference by inferring from smaller randomly sampled subsets of the data on each update. These are called chunks of data, and their size is determined by the chunk size $n_c$. The algorithm will run for a total number of passes through the dataset, defined by $n_p$. We denote the total number of chunks of data processed by N. The algorithm is thoroughly defined as follows:

- **Inputs**
  Event data, and the approximate posteriors $q(t_{j,i}) = \text{Multinomial}(\phi^0_{j,i})$, $q(\omega_j) = \text{Dirichlet}(\gamma^0_{j,t})$, $q(\beta_t) = \text{Dirichlet}(\lambda^0_{t,m})$.

- **Outputs**
  The approximate posteriors $q(t_{j,i}) = \text{Multinomial}(\phi^N_{j,i})$, $q(\omega_j) = \text{Dirichlet}(\gamma^N_{j,t})$, $q(\beta_t) = \text{Dirichlet}(\lambda^N_{t,m})$.

- **Procedure**
  1. Initialise $\lambda^{(n=0)}_{t,m}$.
  2. For chunks $n = 1, \ldots, N$, do:
     *a)* Initialise $\gamma^{(l=0)}_{j,t}$, $\phi^{(l=0)}_{j,i}$.
     *b)* For iterations $l = 1, \ldots, L$ do:
        1) Update $q(\phi_{j,i})$ by iterating through $j$ and $i$ and setting

$$\phi^{(l)}_{j,i}(t) = \frac{e^{\psi\left(\lambda^{(n-1)}_{t,o_{j,i}}\right) - \psi\left(\sum_{m=1}^{M} \lambda^{(n-1)}_{t,m}\right)) + \psi\left(\gamma^{(l-1)}_{j,t}\right) - \psi\left(\sum_{p=1}^{T} \gamma^{(l-1)}_{j,p}\right)}}{\sum_{s=1}^{T} e^{\psi\left(\lambda^{(n-1)}_{s,o_{j,i}}\right) - \psi\left(\sum_{m=1}^{M} \lambda^{(n-1)}_{s,m}\right)) + \psi\left(\gamma^{(l-1)}_{j,s}\right) - \psi\left(\sum_{p=1}^{K} \gamma^{(l-1)}_{j,p}\right)}}. \tag{102}$$

2) Update $q(\omega_j)$ by iterating through $t$ and setting:

$$\gamma_{j,t}^{(l)} = \alpha_t + \sum_{i=1}^{N_j} \phi_{j,i}^{(l)}(t)\,. \tag{103}$$

3) Check for convergence: if the change in $\gamma$ is less than the threshold parameter $\gamma_{\text{thresh}}$, end loop.

4) Set $\phi_{j,i}^{(n)} = \phi_{j,i}^{(L)}$ and $\gamma_{j,t}^{(n)} = \gamma_{j,t}^{(L)}$.

*c*) Update the themes.

1) Update $q(\beta_t)$ by iterating through $t$ and $m$ and setting:

$$\lambda_{t,m}^{(n)} = (1-\delta_n)\lambda_{t,m}^{(n-1)} + \delta_n \left( \eta + \sum_{j=1}^{N_e} \sum_{i=1}^{N_j} \phi_{j,i}^{(n)}(t) \mathbb{I}(o_{j,i} = m) \right). \tag{104}$$

*d*) Evaluate the normalised (per-$o_{j,i}$) ELBO for this chunk of data from the dataset, $\mathcal{L}_n$. This can be used to check for convergence.

- **Return**
  $\gamma_{j,t}^{(N)}$, $\phi_{j,i}^{(N)}$, and $\lambda_{t,m}^{(N)}$.

The algorithm makes use of the hierarchical structure of the model, with local variables $(\omega, t)$ being optimised until convergence before an update on the global variables (themes, $\beta$) is performed. While optimising the local variables the algorithm uses the $(l-1)^{\text{th}}$ approximation of $\gamma_{j,t}$ and the $(n-1)^{\text{th}}$ approximation of $\lambda_{t,m}$ to calculate the $l^{\text{th}}$ approximation of $\phi_{j,i}$, before using the $l^{\text{th}}$ approximation of $\phi_{j,i}$ to calculate the $l^{\text{th}}$ approximation to $\gamma_{j,t}$. Once $L$ updates of this sort have been done, or until convergence has been met according to $\gamma_{\text{thresh}}$, the themes are updated using the local variables obtained at the end of the inner loop.

Two specific notations were introduced when writing the algorithm: the digamma function $\psi(\cdot)$ in Eq. 102, which arises from the expectation of the natural logarithm of the Dirichlet distribution $\psi(x) = d\log[\Gamma(x)]/dx$, (ii) and the indicator function $\mathbb{I}(\cdot)$ in Eq. 104 which is equal to 1 when the equality in the brackets is true, and equals zero when it is not.

It is important to note the key role played by the latent variable $\phi_{j,i}$, which encodes information on the theme assignment per measurement per event. This variable captures the co-occurrences between different measurements in the event sample. For example, if some measurement $m'$ co-occurs with another measurement $m''$ in many events, this information is stored by the $\phi_{j,i}$ variable and through iterative updates these two measurements are more likely to end up with large weights in the same theme distribution. It is through the presence of co-occurring measurements in the data that this algorithm is able to disentangle different underlying physical processes occurring in the events, encoding the new information provided by data on the $\phi_{j,i}$ which are then utilized to update the $\gamma_{j,t}$ and $\lambda_{t,m}$ parameters. The importance of new information is controlled by the prior hyperparameters $\alpha_t$ and $\eta$ that enter Eqs. 103 and 104 respectively. Without these co-occurrences, or a method to utilise them, the best an unsupervised algorithm can do in identifying rare events is to search for outliers in the data. Thus searching for these co-occurrences is essential in extracting a

generative description of the events. We pay particular attention to this in deciding upon a data representation for our benchmark studies in Subsection 6.3.4.

In this Chapter we have used the implementation of the SVI procedure as described above within `gensim` [223], a software package for performing unsupervised semantic modeling of plain text. The parameters of the SVI algorithm are the chunk size $n_c$, the number of iterations L, the threshold $\gamma_{\text{thresh}}$, and the number of passes $n_P$. The learning rate $\delta_n$ is not constant in `gensim` but follows

$$\delta_n = \frac{1}{(\tau_0 + n)^\kappa}, \tag{105}$$

where $\tau_0$ is the offset and $\kappa$ is the decay parameter.

This stochastic inference procedure is proven to converge to a local minimum if $\sum_{n=1}^{\infty} \delta_n = \infty$ and $\sum_{n=1}^{\infty} \delta_n^2 < \infty$, which is guaranteed for $\kappa \in (\frac{1}{2}, 1]$. The convergence of the whole algorithm can be assessed using the ELBO, or equivalently, the perplexity defined as

$$\mathcal{P}_n = 2^{-\mathcal{L}_n}. \tag{106}$$

A lower perplexity means that the ELBO is larger and thus the KL divergence between the posterior and its approximation is smaller. In Section 6.5 we study how the choices of the offset and the chunk size parameters of the algorithm affect the convergence and performance of the models as well as their final perplexity.

With the posterior distributions at hand we can perform a fully bayesian treatment where we account for every possible theme distributions and theme weights of individual events. We could maximise the posterior distributions with respect to the variational parameters to obtain the best estimates of the theme parameters and mixing weights for the LDA model. However this is computationally difficult [224]. A good approximation for the theme parameters and mixing weights can instead be obtained by simply taking the expectation values,

$$\hat{\beta}_{t,m} = \mathbb{E}_q[\beta_{t,m}] = \frac{\lambda_{t,m}}{\sum_{m=1}^{M} \lambda_{t,m}}, \tag{107}$$

$$\hat{\omega}_{j,t} = \mathbb{E}_q[\omega_{j,t}] = \frac{\gamma_{j,t}}{\sum_{t=1}^{T} \gamma_{j,t}}. \tag{108}$$

### 6.1.4 *The LDA landscape*

In Subsection 6.1.2 we discussed the importance of the hyper-parameters $\Sigma$ and $\rho$ in defining the LDA model, and their physical importance with regards to the dataset. However in Subsection 6.1.3 we did not update these parameters during the approximate inference procedure. Although they could be updated through the use of appropriate inverse gradients [221], we choose to consider them fixed as hyper-parameters that define a given LDA model. As we vary these hyper-parameters, we have different values of the ELBO and a different outcome for the learned themes. For this reason we refer to a 'landscape' of LDA models defined over the hyper-parameters.

#### *A landscape of classifiers*

The landscape of LDA models provides different learned themes that could be used to classify events as either signal or background with varying degrees of success. We thus say that we have a landscape of LDA classifiers. Essentially, we want to be able

to cluster events into one of two clusters, $\mathcal{C}_1$ or $\mathcal{C}_2$, using the posterior approximation and the estimates of the theme distributions $\hat{\beta}_{t,m} = \hat{\beta}_{t,m}(\Sigma, \rho)$ and the theme weights for each event $\hat{\omega}_{j,t} = \hat{\omega}_{j,t}(\Sigma, \rho)$. The mixed-membership model assumes that each event is already a mixture of two types of underlying themes, so we could simply cluster the events by placing cuts on $\hat{\omega}_{j,1}$ for each event:

$$\begin{aligned} \hat{\omega}_{j,1} > c \;&\Rightarrow\; e_j \in \mathcal{C}_1\,, \\ \hat{\omega}_{j,1} \leqslant c \;&\Rightarrow\; e_j \in \mathcal{C}_2\,. \end{aligned} \tag{109}$$

Classifying events in this way does yield good classification performance, as demonstrated in Ref. [100]. However, we observe that the learned theme weights are obviously strongly correlated with the learned themes, and we can actually construct a likelihood ratio classifier using the learned themes directly:

$$\begin{aligned} L(e_j|\Sigma, \rho) &= \prod_{i=1}^{n_j} \frac{p(o_{j,i}|\beta_2)}{p(o_{j,i}|\beta_1)} \\ &\simeq \prod_{m=1}^{M} \left( \frac{\hat{\beta}_{2,m}}{\hat{\beta}_{1,m}} \right)^{\mathbb{I}(m=o_{j,i})}\,, \end{aligned} \tag{110}$$

where $\mathbb{I}(\cdot)$ is again the indicator function, equal to 1 when the expression in brackets is true and equal to 0 when it is not. With the likelihood ratio we also need to perform a cut in order to cluster the events,

$$\begin{aligned} L(e_j|\Sigma, \rho) \leqslant c \;&\Rightarrow\; e_j \in \mathcal{C}_1\,, \\ L(e_j|\Sigma, \rho) > c \;&\Rightarrow\; e_j \in \mathcal{C}_2\,. \end{aligned} \tag{111}$$

The likelihood ratio is a more widely used classifier in the HEP and ML communities, and in general we find comparable results when using the likelihood ratio rather than the theme weights, with the former being more stable than the latter.

In expressions 110 and 111 we include the dependence of the likelihood-ratio classifier on the shape of the Dirichlet prior through the two free parameters $\Sigma$ and $\rho$. This highlights that the two-theme LDA model leads to a (two-dimensional) *landscape of classifiers* $L(e_j|\Sigma, \rho)$, where different choices for the prior shape leads to different classifiers with varying performance.

*Evaluating the classifier performance*

Following the standard practices, we evaluate how well a particular classification technique performs by using Monte Carlo generated data, for which we know the truth labels. In order to do this, we generate two samples of events, sample 1 and sample 2, and we produce a mixed sample of events from these pure samples. We train an LDA model with $T = 2$ on this mixed sample to extract 2 theme distributions that describe the data. We then use either the extracted theme weights or the likelihood ratio to cluster the events in $\mathcal{C}_1$ and $\mathcal{C}_2$, the goal being clustering events from sample 1 into $\mathcal{C}_1$ and events from sample 2 into $\mathcal{C}_2$. We test how well the algorithm performs using the truth labelled data. We compute the fraction of events from sample 2 correctly assigned to $\mathcal{C}_2$ as a function of the cut, $\varepsilon_2(c)$ and analogously we compute the fraction of events from sample 1 incorrectly assigned to $\mathcal{C}_2$ as a function of the cut, $\varepsilon_1(c)$. The Receiver-Operating-Characteristic (ROC) curve is then defined as the curve tracing the

true positive rate as a function of the false positive rate $\varepsilon_2(\varepsilon_1)$. The two measures we use to evaluate the performance of the LDA models we have trained are

1. Area Under Curve (AUC): the integrated area under the ROC curve.
2. Inverse mistag at fixed efficiency: $\varepsilon_1^{-1}(\varepsilon_2 = 0.5)$.

The AUC is a useful statistic when we are interested in the general performance of the classifier. However, when the experimental analysis is focused on identifying rare signals in a sample of events the AUC statistic is not always the most relevant indicator of performance. In fact, what is required is a statistic which demonstrates a strong rejection of background events coinciding with the acceptance of a moderately large number of signal events. This is captured by the inverse mistag at fixed efficiency.

*Model selection with perplexity*

Different classification tasks, determined by the data composition and complexity, require different Dirichlet prior shapes and thus different hyper-parameters. For instance, we could intend to extract rare 'anomalous' events from the data. We would then like to classify events into two clusters: one 'normal' cluster $\mathcal{C}_1$ capturing the bulk of events (e.g. SM events), and one 'anomalous' cluster $\mathcal{C}_2$ comprised of a rare events with anomalous features (e.g. BSM). Such a classification is well described by a two-theme LDA with a very asymmetric prior, accomplished by taking $\rho \ll 1$. As we vary the hyper-parameters the performance of the classifiers on the classification task is thus expected to vary greatly. Therefore we need an unsupervised method for determining a near-optimal choice of the hyper-parameters. Here we propose to do this using the model perplexity defined in Eq. 106. This is an intuitive choice, since finding the hyper-parameters which maximise the perplexity is equivalent to finding the hyper-parameters which maximise the ELBO.

In Section 6.4 we study this strategy in detail by performing systematic scans over the hyper-parameter space and correlating the relevant classification performance measures with LDA model perplexity.

## 6.2 LEARNING LATENT JET SUBSTRUCTURE

In Section 6.1 we have introduced probabilistic generative generative models as a tool for analysing experimental data with an emphasis in extracting rare signals from a dataset. As an example of how this works in practice, we apply LDA to the analysis of di-jet events. In this Section we explain how to represent di-jet events in terms of a sequence of exchangeable measurements $o_{j,i}$, discuss how the mixed-membership model is well suited to describe them and consider possible $o_{j,i}$ representations and bases.

### 6.2.1 *Jet de-clustering and substructure observables*

When coloured particles are produced at high energy colliders the subsequent QCD showering, fragmentation, and hadronization results in many hadrons in the final state. If the transverse momentum of the initial particle is large enough, all of these final-state hadrons will be registered by the detector within a single localised region in $(\eta, \phi)$ space. These clusters of hadrons are referred to as jets, and there have

been many different clustering techniques developed to define jets based on the four-momenta of the constituent hadrons. Of these different techniques, the sequential recombination schemes [181, 225–228] have become the standard algorithms for jet clustering. When applied to data collected for a single collider event, the algorithm can reduce the complexity of the data to a handful of jets, each representing a final state of some high-energy parton produced in the hard collision. In order to arrive at a single clustered jet from hundreds of hadrons, the sequential recombination scheme goes through a set of pairwise intermediate clusterings in which the four-momenta of two subjets are combined to form a larger subjet. De-clustering the jet and analysing these individual splittings can provide crucial information into the physical processes taking place during the event. For example, if the initial particle is a top quark with a large transverse momentum, the resulting jet will contain splittings that describe the decay of the top quark into a bottom quark and a $W$ boson, and splittings describing the decay of the $W$ boson. These features are readily exploited by existing traditional top taggers, e.g. Ref. [229].

To analyse di-jet events with the probabilistic generative models outlined in Section 6.1, we extract measurements from each of these splittings in the de-clustering procedure. At each splitting we construct a number of observables using the four-momenta of the subjet being de-clustered ($j_0$), and the two subjets resulting from the de-clustering ($j_1$ and $j_2$). The process for doing this is straight-forward but we must decide on a fixed set of observables to use throughout. The choice of observables is discussed in detail in Section 6.3.

Once we have collected a set of measurements at a splitting, we must then bin their values. Although this is justified by the detector resolution, the reason for binning is to provide a reasonable vocabulary for the algorithm to handle in a finite sample. The relationship between the size of the observable bins and the algorithm performance is discussed in detail in Subsection 6.3.4. One of these binned lists of observables is what we refer to as a measurement $o_{j,i}$ in the probabilistic model. Because of the binning, there will be a finite (although possibly very large) number of values that each $o_{j,i}$ can take.

In addition to the kinematical observables at each splitting we add one more categorical observable, a label identifying which jet the splitting belongs to. With these methods we are describing the whole event rather than a single jet, so the information to which jet a splitting belongs is important to properly characterise the whole event.

Of course, including measurements from all splittings in the jet clustering history is not necessary, and would hinder the VI algorithm in extracting themes relevant for describing a potential signal. Thus we need to impose cuts such that most of the splittings that are irrelevant to uncovering rare signal events are removed. For example, a simple cut on subjet masses removing splittings of subjets with $m_0 < m_{cut}$ could remove many of the soft emissions occurring near the end of the QCD showering process.

The whole process, starting from the raw event data, can be described as follows:

1. Cluster the event with a large jet radius, and keep only the two hardest jets.
2. De-cluster each jet, extracting a list of measurements at each splitting.
3. Bin the measurements from each splitting appropriately, and assign a label identifying which jet the splitting belongs to.
4. Apply kinematical cuts on the splittings.

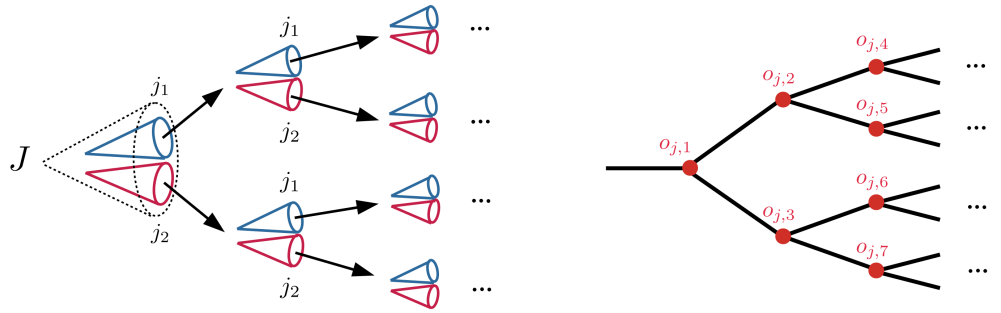



Figure 46: In the left plot we show schematically how the sequential unclustering algorithm proceeds, with the whole jet J being repeatedly separated into two subjets ($j_1$ and $j_2$ with $m_{j_1} > m_{j_2}$). In the right plot we show how the feature representation of the data maps onto this unclustering, with each $o_{j,i}$ being mapped to a node in the unclustering tree. Note that the ordering of these $o_{j,i}$ terms within a single jet does not matter.

An event is then described by an ordered sequence of $o_{j,i}$ each representing a splitting, where each $o_{j,i}$ consists of a list of binned measurements and a label identifying which jet the splitting occurred in. A graphical representing can be seen in Fig. 46.

It is important to highlight that this method, and the model in general, does not rely on any specific clustering scheme and that any set of measurements which describe substructure kinematics of the jets could be used.

### 6.2.2 *Probabilistic models of jet substructure*

At the core of the probabilistic models discussed in Section 6.1 is de Finetti's theorem. Under the assumption that the measurements $o_{j,i}$ used to describe the events are exchangeable, this theorem allows us to derive, based on additional modeling assumptions, the different latent structures in mixture models and mixed-membership models. Constructing the $o_{j,i}$ variables for jet substructure as described in the previous Section is in line with the exchangeability assumption. Sequential jet clustering algorithms do impose an ordering on the splittings due to the pairwise nature of the algorithms and the procedure through which the next subjets to be clustered are selected. However it is the kinematical properties of the splittings that carry most of the interesting physical information, not the order in which they occur, as shown in Ref. [230]. We thus forget the splitting order and focus on the co-occurrences of interesting properties to characterize the relevant physical processes. Thus, the latent themes in both the mixture and mixed-membership models for di-jet events are probability distributions over the space of possible splittings (de-clusterings) that can occur within the two leading jets.

The generative processes for the mixture model and mixed-membership (LDA) model are of course different. In a mixture model a theme would ideally be associated to the specific (hard) partons produced in the collision. Each splitting in an event described by a mixture model is sampled from just one theme, therefore this theme must represent all of the physical processes occurring within the jets produced within that event. In a mixed-membership model however, different themes can be associated to different physical processes occurring within the jets of a single event. Each event in a mixed-membership model is composed of a mixture of themes, just

as there are mixtures of different physical processes occurring within each event. The theme proportions for each event are selected individually from a prior distribution, whose parameters are important in the modeling. Measurements in each event are 'generated' by first sampling theme proportions from the prior, then for each splitting $o_{j,i}$ a theme is drawn from the theme proportions, and a splitting is sampled from that theme.

As an example consider modeling a mixed sample of events consisting of QCD di-jet events ($pp \to jj$), and top quark pair-production events ($pp \to t\bar{t} \to (W^+b)(W^-\bar{b}) \to jj$) where the top quarks are boosted enough such that the decay products of a single top are clustered into a single jet. The splittings within a QCD jet will be predominantly soft with the number of splittings at higher $k_T$ being monotonically suppressed. For the top jets the decay chain also involves many coloured particles (the top, the bottom, the decay products of the $W$ boson), therefore there will be many, predominantly soft, gluon emissions occurring within the top jets as well. However there will always also be a few hard splittings corresponding to the decay of the top quark to the bottom quark and $W$ boson, and the decay of the $W$ boson to light quarks. Using a two-theme mixture to model this event sample would ideally lead to one theme describing all the splittings within QCD jets, and one theme describing both the hard (decay) splittings and soft (QCD) splittings within the top jets. With a mixed-membership model on the other hand, the soft splittings occurring within both the QCD and top jets can be modeled by one theme, with the other theme describing just the hard splittings related to the decay dynamics inside top jets. This seems like a natural setting in which to search for rare BSM signals in di-jets at high-energy colliders.

### 6.2.3 *Choosing a data representation for the jet substructure*

The discussion so far has not been specific about the observables to be measured at each $j_0 \to j_1 j_2$ splitting in the jets. In this Subsection we discuss and justify two bases of observables, with each basis using a different cut to determine which splittings are included in the analysis. Moreover, in the end we only use a subset of the observables from each basis, as explained in more detail in Section 6.3.
The first choice is what we refer to as the mass basis, see Refs. [229, 231]:

$$\textit{mass-basis}\text{: } \{m_0, \tfrac{m_1}{m_0}, \tfrac{m_2}{m_1}, \tfrac{k_T}{m_0}, \cos\theta\} \text{ where } m_1 > m_2.$$

These are the mass of the (mother) (sub)jet being de-clustered, the mother/daughter subjet mass drop, the daughter subjets' mass ratio, the $k_T$ distance between the daughter subjets defined in the usual way as $k_T = p_{T,2}\Delta$, where $\Delta^2 = (y_2 - y_1)^2 + (\phi_2 - \phi_1)^2$, and the helicity angle between the mother (sub)jet and the daughter subjets as defined in Refs. [232, 233]. In this basis we only include splittings from the jets in which the subjet being de-clustered has a mass $m_0 > 30$ GeV.
The second choice is what we refer to as the Lund basis, see Ref. [230]:

$$\textit{Lund-basis}\text{: } \{m_0, \log R/\Delta, \log k_T, \log R/\kappa, z, \psi\},$$

where $R$ is the jet radius, $z = p_{T,2}/(p_{T,1} + p_{T,2})$, $\kappa = z\Delta$, $\psi = \tan^{-1}(y_2 - y_1)/(\phi_2 - \phi_1)$, and $p_{T,1} > p_{T,2}$. In this basis we only include splittings from the jets which lie on the primary Lund plane. This plane is defined as the path through the clustering history starting from the clustered jet and continually moving through the pairwise splittings to the subjet with the largest $p_T$ until the end of the clustering history. One advantage of the primary Lund plane compared to the mass-basis is that it offers a

clearer interpretation in terms of hard vs. soft (and perturbative vs. non-perturbative) splittings, see Ref. [230] for details.

We emphasise that these two bases do not just differ in the observables, but the different cuts make a considerable difference in the splittings which are used for the description of the jets. In Subsection 6.3.4 we explore how some features of the dataset change as we vary the binning used for these observables. Here we only specify the default bin sizes which we use in Section 6.4: for the mass-basis observables we bin the measurements in intervals of $\{10\,\text{GeV}, 0.05, 0.05, 0.05, 0.1\}$, while for the Lund basis we use $\{10\,\text{GeV}, 0.2, 0.2, 0.05, 0.2, 0.2\}$.

The last thing to discuss in terms of the data representation are the jet labels. In Section 6.2.1 we discussed the importance of including jet labels to differentiate between splittings occurring in the two jets, however we did not specify how these jets should be labelled. Naively, because we select the jets according to $p_T$, one might choose to also label the jets in the same way with $J_1$ being the jet leading in $p_T$ and $J_2$ being the jet subleading in $p_T$, thus labelling with the condition that $p_{T,J_1} > p_{T,J_2}$. However this is not suitable in practice. In the top quark pair production example discussed above, the ordering of the jet labels is not so important, since both jets in the event are top jets and have the same decay structure. However, one can easily imagine different signals that are not so simple. In many cases, including the BSM example studied in this Chapter, the two jets in the final state will have been seeded by two different particles of different mass and thus they will both have distinctly different decay dynamics. Being able to differentiate between these different structures is not just important for classification, but is also important for a physical interpretation of the themes learned through the VI algorithm. Therefore in the case where the signal events contain two different jets, we would like to be able to associate the $(J_1, J_2)$ labels with splittings from one jet or the other, consistently across the whole sample. This will not happen if we label the jets by their $p_T$, instead the best way to do this is by labelling the jets according to their jet mass $m_J$, such that $m_1 > m_2$.

## 6.3 SET-UP AND BENCHMARKS

In this Section we detail the specific set-up and benchmark scenarios we consider, including the appropriate choice of observable basis and binnings.

### 6.3.1 *Algorithm set-up*

As discussed in Subsection 6.1.3, there are a number of parameters that determine how the VI algorithm is implemented. In our benchmark examples we use the following choices which produce robust results across a wide range of scenarios: passes $n_p = 200$, chunk size $n_c = 10^4$, iterations $L = 100$, offset $\tau_0 = 1000$, $\gamma_{\text{thresh}} = 10^{-8}$, decay $\kappa = 0.5$. These choices are justified in Section 6.5 where we discuss in particular how changing the chunk size and the offset affects the convergence of the algorithm and the performance of the classifier.

### 6.3.2 *Benchmark di-jet events*

We perform our analysis using two benchmark scenarios, (i) boosted top quark pair-production $pp \to t\bar{t} \to b\bar{b}W^+W^-$, and (ii) a hypothetical 3 TeV vector $W'$ plus a 400

GeV scalar $\phi$ model, with the dominant production and decay chain $pp \to W' \to W(\phi \to WW)$. Since our model is built to learn from jet substructure, in both cases we consider only the hadronic final states of the $W$ bosons. Consequently, the main background process in both scenarios is the QCD di-jet production.

All event samples were generated using `MadGraph5_aMC@NLO` [12] interfaced with `Pythia 8` [16] for showering and hadronization, and `FastJet 3.4.1` [234] for jet clustering. The events were generated at a collision energy of 13 TeV and the jets were clustered using the CA algorithm [227, 228] with $R = 1.5$. No jet grooming was performed. Finally, for $t\bar{t}$ ($W'$), jets with $p_T < 300$ (400) GeV were discarded. The detector effects were not simulated, however we checked that the effects of subcluster energy smearing consistent with the `Delphes 3` [18] simulation of the ATLAS detector had no significant effect on the results.

*Boosted top quark pair-production*

In recent years the $pp \to t\bar{t} \to b\bar{b}W^+W^-$ process has become a standard benchmark for supervised machine learning applications to particle physics [235]. Although there is no need for an unsupervised top tagging algorithm, it is a nice example that demonstrates the power of these techniques by applying them to a well measured and understood physical process. We plot the pure signal ($t\bar{t}$ jets) and background (QCD di-jets) samples in the $(m_0, m_1/m_0)$ and $(\log R/\Delta, \log k_T)$ planes, in Figs. 47 and 48 respectively.

We see in Fig. 47 that the hard splittings corresponding to the decay of the top quark to the $W$ boson and the decay of the $W$ boson to light jets are clearly discernible. The top quark decay is indicated by the two clusters (overdensities) of measurements at $m_0 \simeq 175$ GeV, with the cluster at $m_1/m_0 \simeq 1$ being due to the clustering of light radiation around the subjet containing all of the top quark decay products, while the cluster at $m_1/m_0 \simeq m_W/m_t$ corresponds to the splitting that separates the bottom and $W$ subjets from within the top jet. The decay of the $W$ boson is indicated by the two clusters at $m_0 \simeq 80$ GeV. Again the cluster at $m_1/m_0 \simeq 1$ is due to the clustering of soft radiation around the subjet containing the $W$ boson decay products, while the cluster at lower mass drop shows splittings that separate the decay products of the $W$ boson. The fact that this cluster is at $m_1/m_0 \simeq 0.2$ does not indicate that the $W$ boson is decaying to a state of mass $\simeq 0.2 m_W$. This feature is an artifact of the definition of mass drop $m_1/m_0$ with $m_1 > m_2$ ordering. Because we take $m_1$ to be the heaviest of the subjets in the splitting, the distribution of the mass drop is skewed away from zero. If we instead had plotted $m_2/m_0$ we would see that this cluster is pushed towards $m_2/m_0 \simeq 0$. The $(m_0, m_1/m_0)$ distribution for the background QCD jets is smooth and monotonically decaying at large $m_0$ and small $m_1/m_0$.

In Fig. 48 we see that the splittings corresponding to the hard decays of the top quark and the $W$ boson are indicated by the two overlapping clusters at $\log k_T \simeq 5$ and $\log R/\Delta \simeq 1$. Apart from the obvious difference in choice of observables here, one should also keep in mind that the actual splittings which pass the cuts here are different than those that pass the cuts in Fig. 47 as detailed in Subsection 6.2.3. This choice leads to a larger overlap between the background and signal distributions, as seen by the stream of splittings at low $\log k_T$. However, there is still a good separation between the features that distinguish the $t\bar{t}$ jets from the QCD background jets.

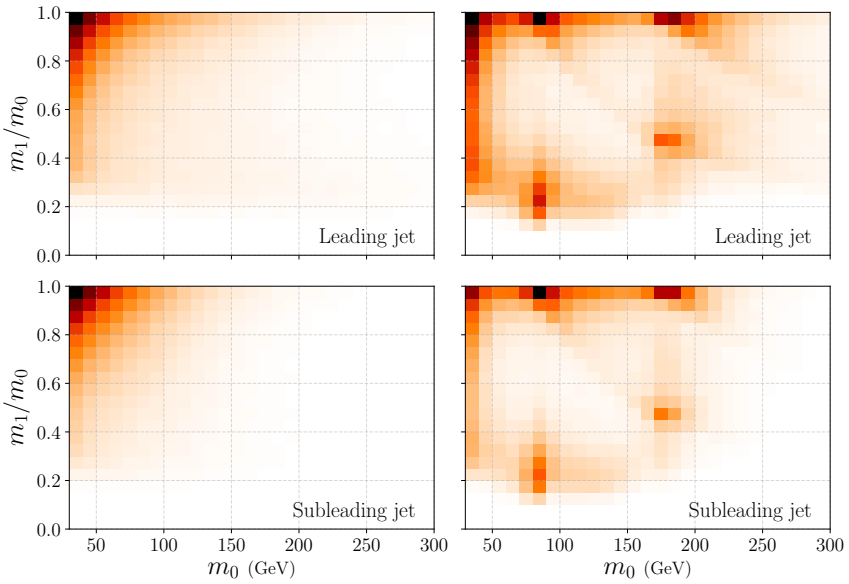



Figure 47: Distributions of QCD (left) and $t\bar{t}$ (right) di-jet events in the ($m_0$, $m_1/m_0$) plane. See text for details.

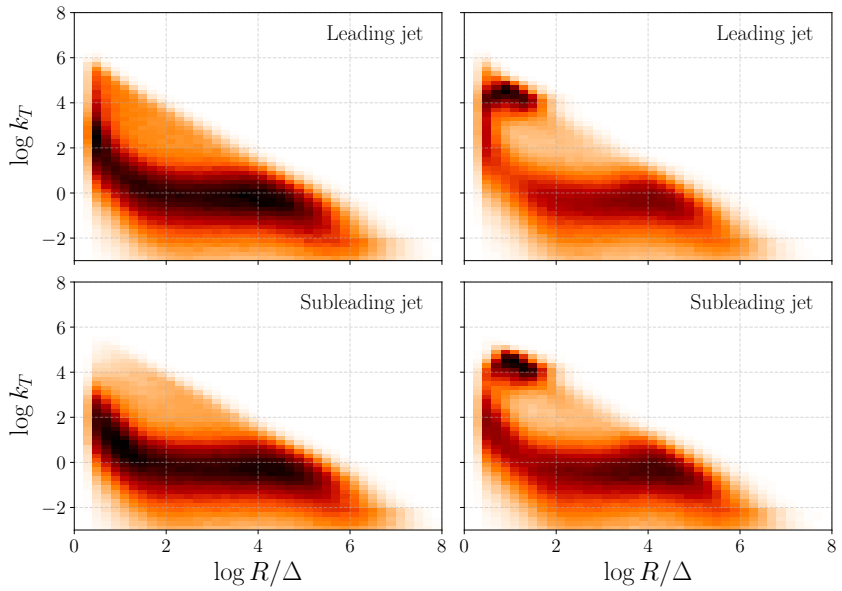



Figure 48: Distributions of QCD (left) and $t\bar{t}$ (right) di-jet events in the ($\log k_T$, $\log R/\Delta$) plane. See text for details.

*A 3 TeV $W'$ model with a 400 GeV scalar*

The second benchmark is an example of a BSM signature which could be searched for at high-energy colliders using these techniques. The BSM process is the production of a 3 TeV $W'$ boson at a collision energy of 13 TeV, which decays to a SM $W$ boson and a 400 GeV BSM scalar boson $\phi$. The scalar boson $\phi$ then subsequently decays to two SM $W$ bosons. The model has been previously introduced and studied in Refs. [74, 81, 87]. The mass difference between the $W'$ and its decay products implies that the constituents from the scalar boson and the $W$ will be clustered into a pair of boosted jets, making the jet substructure an important tool for any analysis of these events.

This was first studied in Ref. [81] and used as a benchmark for the unsupervised CWoLa search technique in Ref. [74] and for the LDA implementation in Ref. [100]. For this benchmark, in addition to the $p_T$ cut at 400 GeV, events were selected in the di-jet invariant mass window $[2700, 3300]$ GeV to encapsulate the peak in the production cross-section of the $W'$ boson.

In Figs. 49 and 50 we plot the pure signal (hadronic $W'$ final states) and background (QCD di-jets) samples in the $(m_0, m_1/m_0)$ and $(\log R/\Delta, \log k_T)$ planes, respectively. The origin of the features in these plots is completely analogous to those of $t\bar{t}$ in Figs. 47 and 48. One important difference to note is that the plots for the leading and subleading signal jets are different, while for $t\bar{t}$ they are equivalent. This is obviously because in this case our signal consists of two jets of different origin, highlighting the importance of including labels for the jets in our representation of the measurements in LDA to properly characterise the signal from the posterior theme distributions.

In Fig. 49 we clearly see the clusters corresponding to the decays of the scalar boson in the leading jet at $m_0 \simeq 400$ GeV. Of the two clusters at $m_0 \simeq 400$ GeV, the cluster at mass drop $m_1/m_0 \simeq 1$ again corresponds to the clustering of soft radiation around the subjet containing all of the scalar boson decay products, while the cluster at $m_1/m_0 \simeq m_\phi/m_{W'}$ corresponds to splittings that separate the SM $W$ boson subjets from within the $\phi$ jet. The clusters corresponding to the decay of the SM $W$ bosons at $m_0 \simeq 80$ GeV have the exact same features as those in the $t\bar{t}$ case.

The distributions for the background jets in Fig. 50 are similar to the distributions for the background jets in Fig. 48, as expected. Interestingly, the distributions for the $W'$ and $t\bar{t}$ signal jets in Figs. 50 and 48 are more similar than they are in Figs. 49 and 47. This is because the observables in the former case measure the $k_T$ and angular separation, rather than the masses of the (sub)jets in the splittings. Also, the observables are now both binned and displayed on a logarithmic scale, making any differences at large $k_T$ less pronounced. One obvious difference between the $W'$ and $t\bar{t}$ distributions in Figs. 50 and 48 is that the clusters associated with the different hard decays are more distinguishable from each other in the $W'$ case than in the $t\bar{t}$ case. This is primarily because the mass difference between the scalar boson $\phi$ and the SM $W$ bosons is much larger than the mass difference between the top quark and the SM $W$ bosons. Another difference is in the amount of soft radiation in the $t\bar{t}$ jets and the $W'$ jets, which is due to the top quark carrying colour charge and the $\phi$ boson being colour-neutral. The similarities in the two distributions do however suggest that any classifier selecting events with splittings in the large $k_T$ region may work reasonably well as a generic anti-QCD tagger.

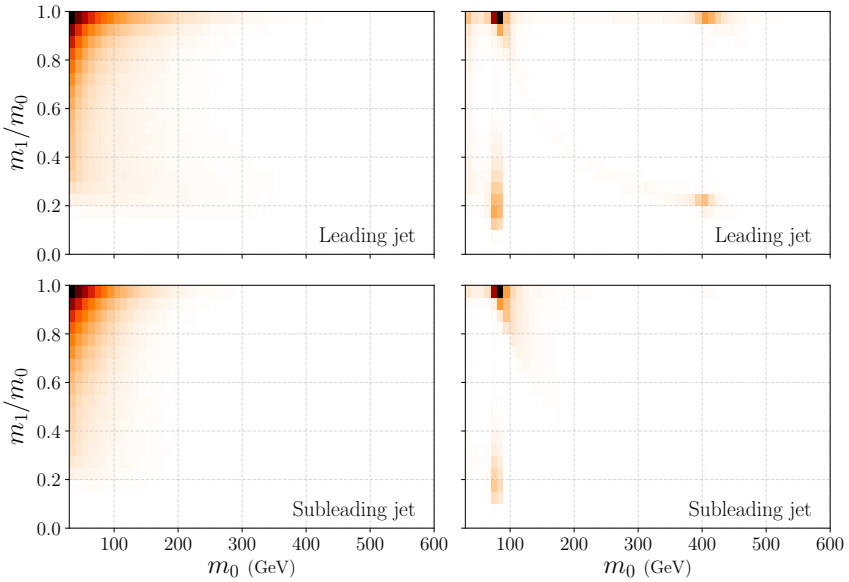



Figure 49: Distributions of QCD (left) and $W'$ (right) di-jet events in the ($m_0$, $m_1/m_0$) plane. See text for details.

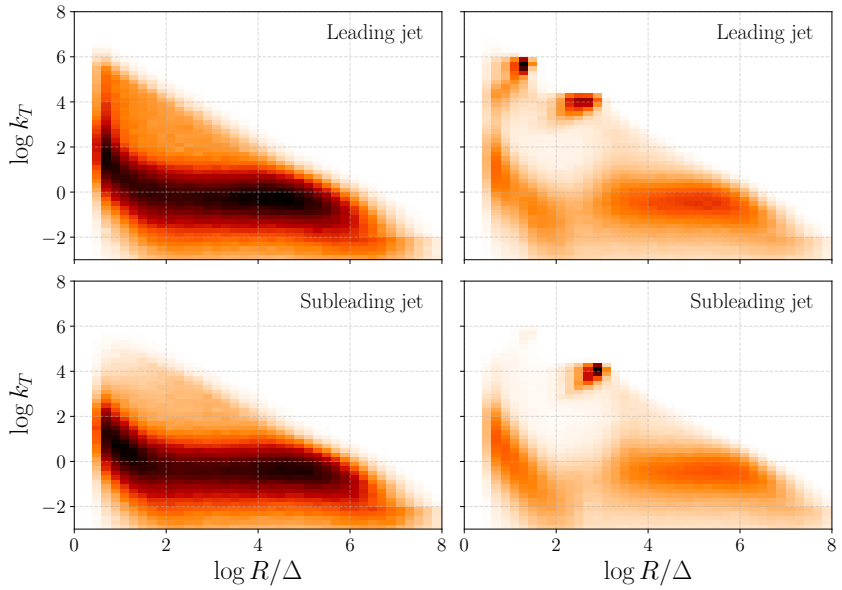



Figure 50: Distributions of QCD (left) and $W'$ (right) di-jet events in the ($\log k_T$, $\log R/\Delta$) plane. See text for details.

### 6.3.3 *Comparing classification power of different observables*

There are many possible choices of observables that we could consider in our analysis of di-jet events using LDA. Although all of the observables discussed in Subsection 6.2.3 carry some ability to distinguish between signal events and QCD background events, some will be more useful than others depending on what the signal process is. In this Section we study the classification power of each of these observables, and some combinations of them, using a simple binned likelihood classifier. To construct the binned likelihood classifier we split our signal and background datasets into 'training' and 'testing' sets. We then compile counts of how often each measurement bin occurs in each of the signal and background training sets, and normalise these to give us a discrete probability distribution for the signal and background samples. For each event in the testing sets we compute the likelihood ratio defined in Eq. 110, except with the $\beta$'s replaced with the binned likelihood multinomials. The results are summarised in Fig. 51.

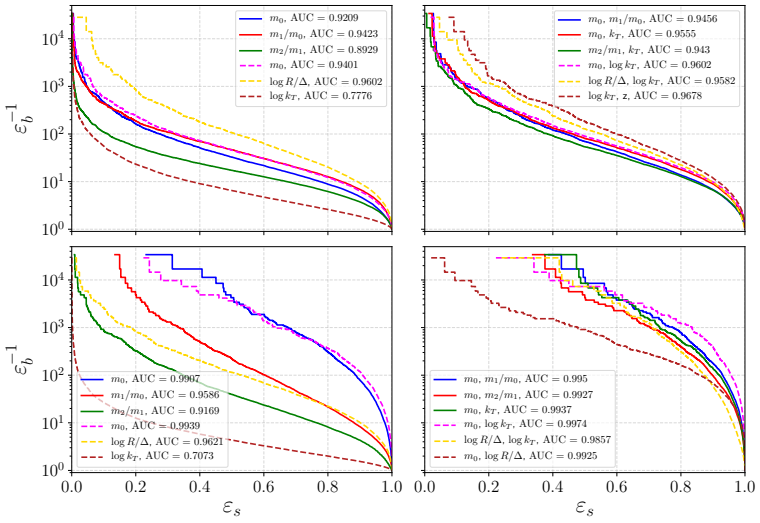



Figure 51: Classification power of individual observables (left column) and pairs of observables (right column) for both $t\bar{t}$ (top row) and $W'$ (bottom row) signals versus QCD. We consider both mass basis observables (solid lines), and observables in the primary Lund plane (dashed lines).

In the first row we show the classification performance of the observables for the $t\bar{t}$ sample. The best performing individual observables are $\log R/\Delta$ and $m_0$ from the Lund basis. Note that $m_0$ appears twice, once in the mass basis and once in the Lund basis. The difference in classification power here comes only from the different cuts performed on the dataset as explained in Subsection 6.2.3. In combining observables we see that the best performing pair of observables in the Lund basis are $\log k_T$ and $z$. In the mass basis the best performing pair are $m_0$ and $k_T$, however the differences between this pair and others are minuscule.

In the second row we show the analogous plots demonstrating the classification power of the observables for the $W'$ sample. The results here are different than for $t\bar{t}$, which is not surprising since not only are the masses of the particles produced in the collision different, but also the top quarks produce a very different radiation pattern than the $W'$ decay products due to the colour and spin differences(two colour and spin 1/2 particles compared against colourless $W$ with spin 1 and $\phi$ with spin 0).

The best performing individual observable here is the subjet mass $m_0$ in both bases, mass and Lund. In combining observables we find that the best performing pair of observables in the Lund basis are $m_0$ and $\log k_T$, while in the mass basis the best performing pair are $m_0$ and $m_1/m_0$. Again the differences between these pairs and some of the others are very small.

We do not study combinations of more than two observables because we find that adding more observables to the best performing pairs does not provide any appreciable difference in classification power of the binned likelihood. The observables are in general better at classifying $W'$ events than $t\bar{t}$ events. This is simply because the $W'$ signal contains splittings that are very rare in QCD background events, in particular rarer than the splittings in $t\bar{t}$ events. Interestingly, we find in Subsection 6.3.4 that the observables which provide the best performance with the supervised binned likelihood classifier are not necessarily the best to use in an unsupervised analysis based on LDA and VI, which crucially depends on patterns of co-occurrence of two or more measurements within the same event. Therefore in the unsupervised analyses we focus solely on two pairs of observables (i) $(m_0, m_1/m_0)$, (ii) $(\log k_T, \log R/\Delta)$, based on their robust performance, good interpretability and established usage in jet classification tasks. Note that while the classification power of any combination of observables in the supervised binned likelihood does not necessarily indicate the best choice to use in any given analysis, it does represent an upper bound on the classification power of any unsupervised classifier based on the corresponding themes of these same binned observables, such as LDA. This is due to the Neyman-Pierson lemma [236].

### 6.3.4 *Measurement co-occurrences*

When introducing LDA and VI in Section 6.1 we encountered the importance of measurement co-occurrence in individual events. Certain measurements within individual events must exhibit a pattern of co-occurrences in order for the inference algorithm to recognise and extract the corresponding theme distributions. In other words, VI is unable to extract any information from unique measurements $o_{j,i}$ appearing only once in the dataset. In this section we explore how the measurement co-occurrences in a dataset vary with the choice of observables and their binning, highlighting the importance of the data representation in the construction of the unsupervised classification strategy. We demonstrate this on the example of the $W'$ model, having checked that the $t\bar{t}$ example exhibits analogous behaviour.

We quantify the measurement co-occurrences by obtaining the number of unique possible $o_{j,i}$ per bin of one of the observables and dividing this by the total number of measurements in that bin. The lower this 'fraction of unique $o_{j,i}$' is, the stronger the co-occurrences are for that bin and the easier it will be for the VI algorithm to extract themes that accurately describe the underlying structure of the events. In the upper row of Fig. 52 we show how the co-occurrences in the mixed $W'$-QCD sample with observables $(m_0, m_1/m_0, m_2/m_1, k_T, \cos\theta)$ vary per $m_0$ bin.

On the left hand side we do this for a mixed sample of $9\times10^4$ signal and background events with varying S/B, while on the right hand side we do it for varying amounts of pure signal events (100, 500, and 1000). We focus on such small numbers of signal events because we are interested in finding rare signals, in which the co-occurrences will inevitably be less apparent. Conversely, in a sample containing a large fraction of signal events the structure of the signal events would be more easily uncovered due to the strong co-occurrences between the measurements. As expected, the co-

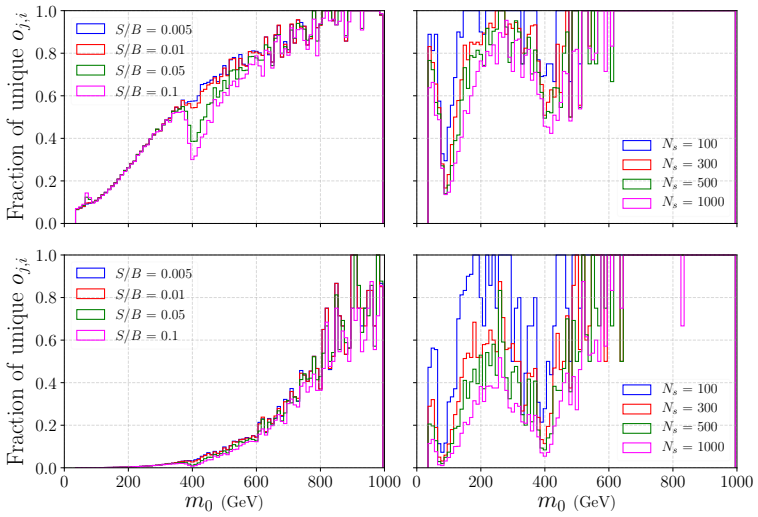



Figure 52: Fraction of unique measurements in $W'$ event samples, using all the mass basis observables ($m_0$, $m_1/m_0$, $m_2/m_1$, $k_T$, $\cos\theta$) (top line) and only the pair ($m_0$, $m_1/m_0$) (bottom line). On the left we show samples of $9\times10^4$ events consisting of different fractions of mixed signal ($W'$) and background (QCD) events, while on the right we show the results for different numbers of pure signal ($W'$) events.

occurrences are strongest at $m_0 \simeq m_W$ and $m_0 \simeq m_\phi$, since the signal events are more likely to contain splittings with these masses as in Fig. 49. However, as discussed in Subsection 6.3.3 including more than two of these observables in the analysis does not significantly increase the classification power of the binned likelihood classifier. At the same time we demonstrate in the second row how restricting the analysis to including just one such pair ($m_0$, $m_1/m_0$) can drastically increase the strength of the co-occurrences in the event sample. This provides further justification for including no more than two observables in the LDA analysis.

In Fig. 53 we display the same information for the Lund observables where we measure the co-occurrences as a function of $\log k_T$. We see again that the co-occurrences are strongest at the points where the signal features are most pronounced (see Fig. 50), and that by restricting the observables used at each splitting we can increase the frequency of these co-occurrences significantly.

Before moving on we examine another handle we have on increasing co-occurrences in the event sample: varying the binning used for each of the observables. To do this we again focus on the $W'$ sample with the pair of ($m_0$, $m_1/m_0$) observables. The results are summarised in Fig. 54. We note that some of the bin sizes used in this plot would be impossible to use in practice due to the finite experimental resolution, however they still serve as useful examples to demonstrate the potential effects of varying bin sizes in the analysis.

In the upper left plot we show the co-occurrences for the whole mixed sample with S/B=5% and four different choices of bin sizes. In each of the other three plots we show the co-occurrences for different numbers of signal events (again 100, 500, and 1000) and varying bin sizes. As expected, larger bin sizes result in stronger co-occurrences, although this effect is subleading with respect to removing observables from the analysis completely. For example, all the studied numbers of signal events show almost the same strength of co-occurrences at $m_0 \simeq m_W$ for all choices of binning. The effect due to different bin sizes is more clearly seen away from these areas of

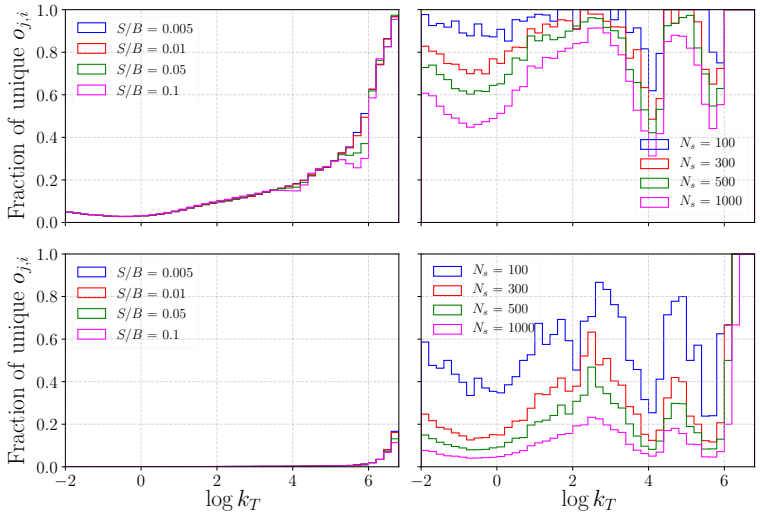



Figure 53: Fraction of unique measurements in $W'$ event samples, using all the Lund basis observables ($m_0$, $\log R/\Delta$, $\log k_T$, $z$, $\log R/\kappa$, $\psi$) (top line) and only the pair ($\log k_T$, $\log R/\Delta$) (bottom line). On the left we show samples of $9\times10^4$ events consisting of different fractions of mixed signal ($W'$) and background (QCD) events, while on the right we show the results for different numbers of pure signal ($W'$) events.

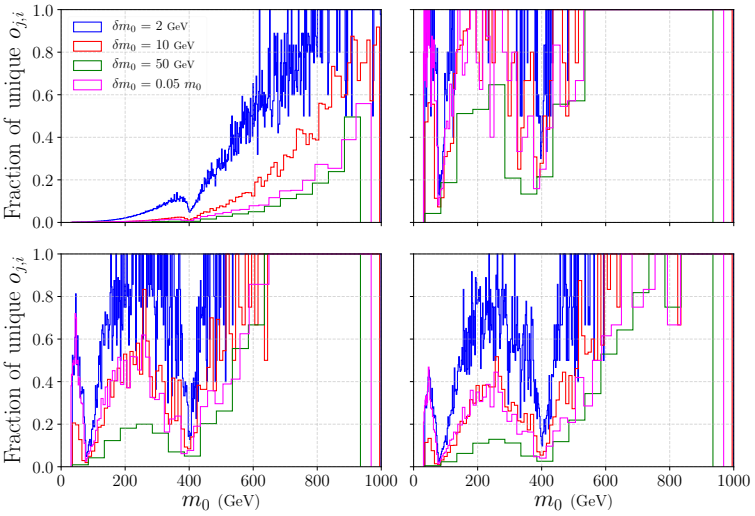



Figure 54: Fraction of unique measurements in $W'$ event samples, using the mass basis observables ($m_0$, $m_1/m_0$) for different choices of $m_0$ binning. From top-left to bottom-right: results for a mixed sample of $9\times10^4$ events with S/B = 5% and four different choices of bin sizes, followed by results for 100, 500 and 1000 pure signal events at various bins sizes of $\delta m_0 = 2$ GeV, 10 GeV 50 GeV, and $0.05 \times m_0$.

strongest co-occurrence, where larger bin sizes result in stronger co-occurrences across the whole $m_0$ range. In particular, this may aid in better modelling of the signal and background distributions away from $m_0 \simeq m_W$ and $m_0 \simeq m_\Phi$. On the other hand, increasing the bin size will also make the signal features less pronounced, potentially reducing the classification power in the same way as a binned likelihood classifier becomes worse and worse approximation to the Neyman-Pearson un-binned likelihood. Therefore there is trade-off here between potential classification power and the ability of VI to extract optimal theme distributions from the data. In particular we find that the constant $\delta m_0 = 10$ GeV bin size provides the best trade-off for the examples stud-

ied here and is also in practice close to the variable binning $\delta m_0 = 0.05 m_0$ mimicking the typical energy resolution of modern particle detector calorimeters.

## 6.4 UNSUPERVISED LEARNING WITH LDA

In this Section we demonstrate the power of the technique outlined in the preceding sections. For this demonstration, we focus on the $W'$ benchmark and the primary Lund basis of observables outlined in Subsections 6.2.3 and 6.3.3. We also show the analogous results using the mass basis of observables and for the $t\bar{t}$ benchmark with both observable bases.

Using the $W'$ benchmark we construct a number of mixtures of background and signal events or mixed event samples. The mixed samples contain $9 \times 10^4$ events with S/B ratios 0.1, 0.05, 0.025, 0.01, and 0.005. We also include a pure background sample (S/B=0) to demonstrate what the output of LDA looks like with no signal events present. The truth level Lund plane distributions for the signal and background events in the $W'$ benchmark are shown in Fig. 50. The background QCD events are described by smooth distributions in which the low $\log k_T$ regions depict the non-perturbative splittings in the QCD jets. For the signal events we see a different behaviour, with both the leading and subleading jets displaying localised high density regions corresponding to the decays of the 400 GeV scalar boson and the $W$ bosons. Identifying these localised high-density regions is crucial to achieving a high-performing classifier.

For each event sample, 6 in total, we train LDA models with different Dirichlet parameters, extract the themes and use them to cluster/classify the events in the sample. We perform an extensive scan over the Dirichlet parameters, scanning over the ($\Sigma$,$\rho$) parameter space in the ranges $0 \leqslant \Sigma \leqslant 3$ and $-3 \leqslant \log_{10} \rho \leqslant 0$ with resolution $\delta\Sigma = 0.1$ and $\delta \log_{10} \rho = 0.1$, respectively. We have then 961 models for each event sample. After training these models, we plot the inverse perplexity ($\mathcal{P}^{-1}$) calculated over the whole sample, as well as the AUC and the inverse mistag at fixed efficiency. In particular, we are interested in how the inverse perplexity, which can be computed from unlabelled data alone, is correlated with the performance of the labelled metrics (AUC and $\epsilon_b^{-1}$ ($\epsilon_s = 0.5$)), which is inaccessible in absence of labelled data. We also consider how both the inverse perplexity and performance behave in different regions of the Dirichlet parameter space, as discussed in Subsection 6.1.2.

The results of the scan are presented in Fig. 55. An immediate result is the strong correlation between the perplexity measure and the performance of the classifier, which is necessary for an unsupervised search to be possible. In both the perplexity and the performance landscape scans we see a ridge-like structure at $\Sigma < 0.5$ and $\rho \to 0$, beyond which the performance of the classifier degrades. In the $\Sigma \to 0$ limit the LDA model becomes more and more like a mixture model rather than a mixed-membership model. In this case it is known that variational inference techniques tend to perform badly when describing datasets whose themes share features [208] as we detailed when describing mixture models. Given that both our signal and background events share similar non-perturbative features, it is expected that we see this degradation in performance at low $\Sigma$. In the $\rho \to 0$ limit we also see a degradation in the perplexity, along with a more subtle degradation in performance. This again is expected, since in the vanishing $\rho$ limit the LDA model effectively consists of just a single theme and is thus less sufficient in describing the data.

One interesting point worth mentioning here is that even in the $S/B = 0$ limit the classifiers retain some classification power. This is entirely due to an anti-QCD

tagging effect, where the LDA model learns the QCD theme distribution and one other distribution which is typically a mixture of QCD features and random noise. Then through the likelihood ratio these themes will result in a classifier which tends to assign lower values to events which are QCD-like. We also note that in this case the performance is more uniform in $\Sigma$ and $\rho$ than for the large S/B cases, however the ridge-like structure still persists because both at low $\Sigma$ and at large $\rho$ the variational inference algorithm tends to split the QCD features between the two themes. This results in a classifier which does not perform well as at anti-QCD tagging.

The green star in Fig. 55 indicates the point of minimum perplexity for the S/B=5% case, which would be selected as the optimal model in a realistic unsupervised analysis. The inferred themes of this model are shown in Fig. 56. Comparing these themes with the truth-level distributions shown in Fig. 50 we see that the typical signal features are well distinguishable in one of the themes (theme 2). The two clusters in the leading jet distribution of this theme correspond to the decays of the $\phi$ boson and the $W$ boson, with the single cluster in the subleading jet corresponding to the decay of the $W$ boson.

As in the mass basis example for boosted $t\bar{t}$ in Fig. 59, we observe some notable differences when comparing the two themes to pure signal and background distributions, especially in the soft ($k_T \to 0$) and collinear ($\Delta \to 0$) regions of the Lund plane. Theme 2 predominantly captures the hard splittings associated with the massive resonance decays, while the softer splittings are predominantly captured by theme 1. Interestingly, it seems that the algorithm in this case picked up some distinguishable features of the signal (a deficit below the $W$ peak) even in the non-perturbative (low $k_T$) regime. We warn however that these effects are very subtle and the least robust, since they vary considerably with the model priors.

The LDA hyperparameter scan for the $W'$ benchmark in the mass basis can be found in Fig. 57. The scan for the $t\bar{t}$ benchmark in the mass basis in Fig. 58 and in the primary Lund basis in Fig. 60. We show the inferred themes for the maximum perplexity model for the $t\bar{t}$ S/B = 0.1 in the mass basis in Fig. 59.

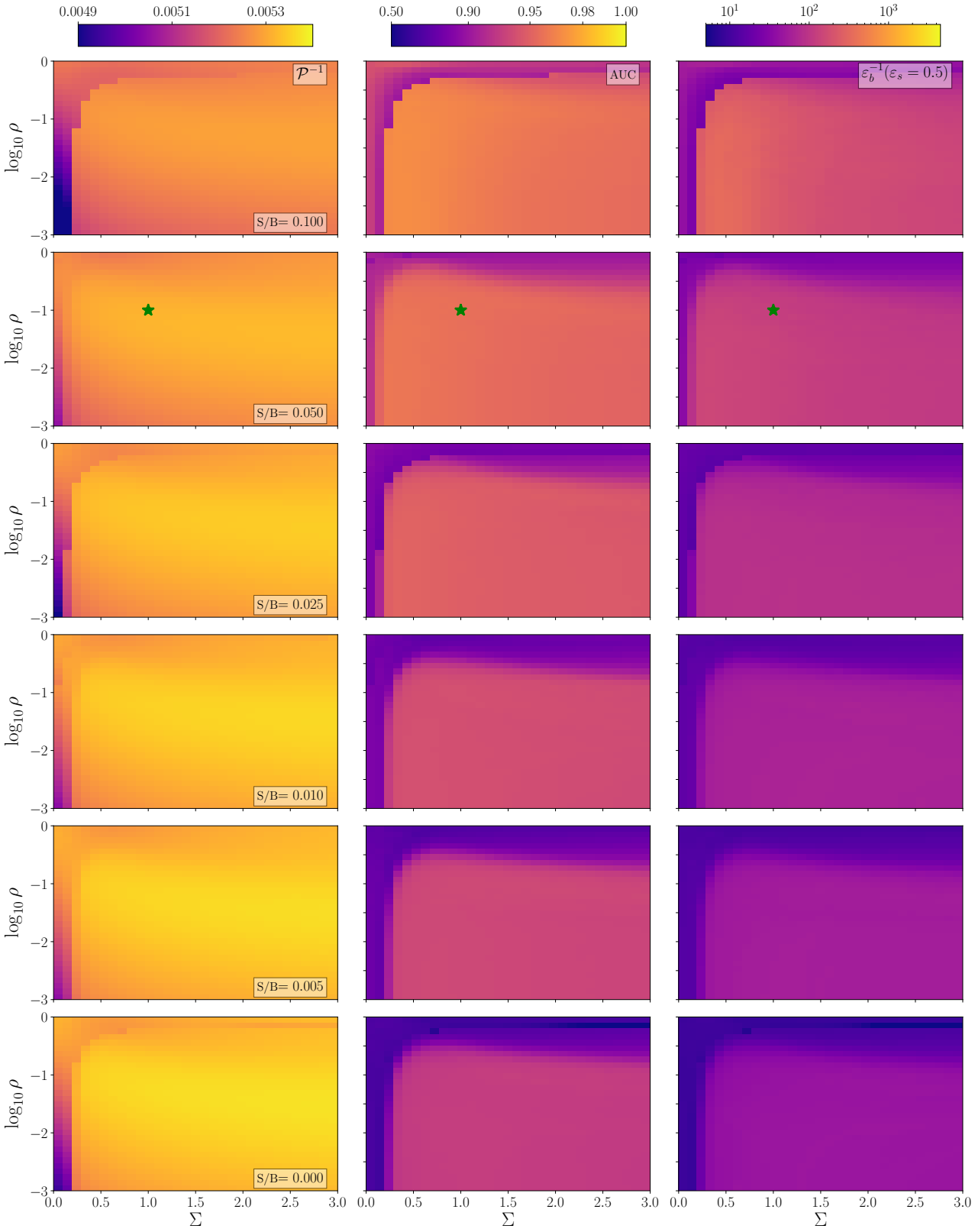



Figure 55: Results of LDA models in the $(\Sigma, \rho)$ parameter-space trained on samples of mixed $W'$ and QCD events using Lund basis observables $\log k_T$ and $\log R/\Delta$, with different S/B ratios (one per row). Each row contains plots of perplexity, AUC, and inverse mis-tag rate at fixed efficiency (see text for details). The green star indicates the model used to plot the theme distributions in Fig. 56.

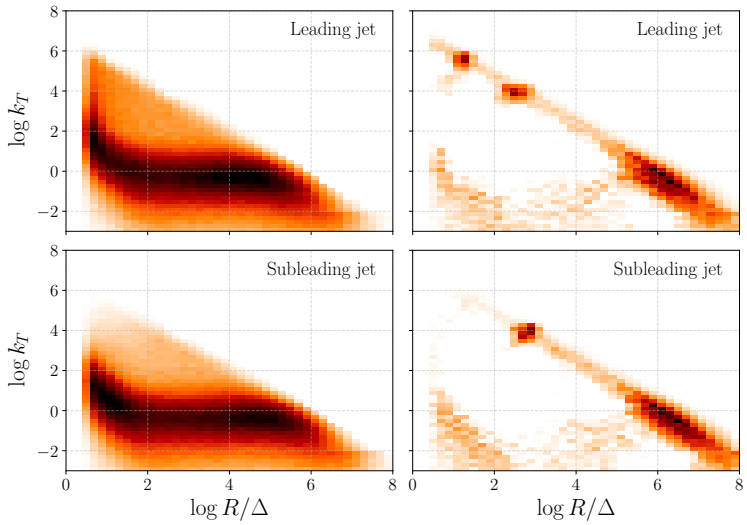



Figure 56: The LDA extracted theme 1 (left) and theme 2 (right) distributions for the leading (upper plots) and subleading (lower plots) jets obtained on a mixed $W'$ / QCD sample with S/B = 0.05, where only the $\log k_T$ and $\log R/\Delta$ observables were used. The results are obtained using the model that yields the lowest perplexity. See text for details.

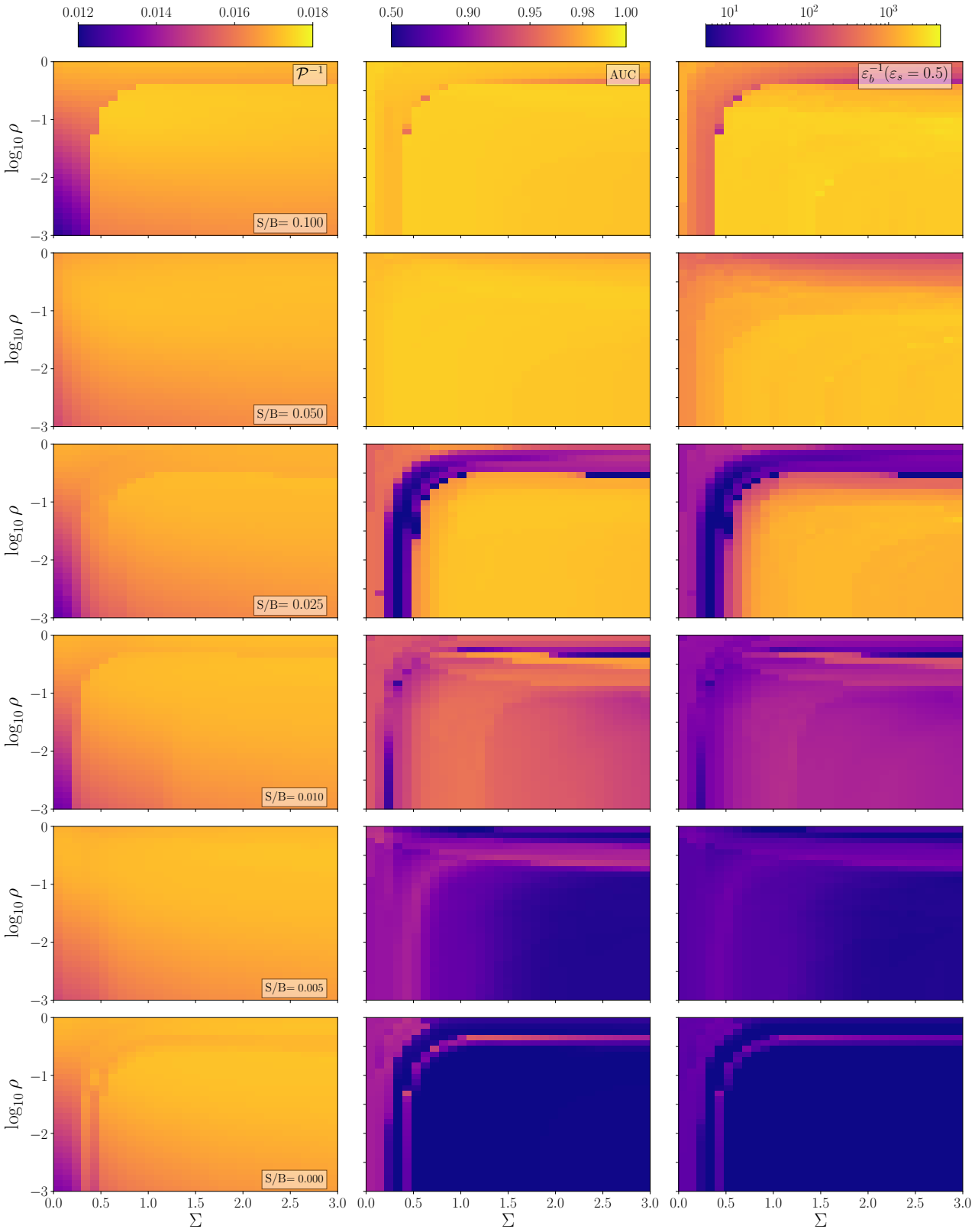



Figure 57: Results of LDA models in the $(\Sigma, \rho)$ parameter-space trained on samples of mixed $W'$ and QCD events using mass basis observables $m_0$ and $m_1/m_0$, with different S/B ratios (one per row). Each row contains plots of perplexity, AUC, and inverse mis-tag rate at fixed efficiency. See text for details.

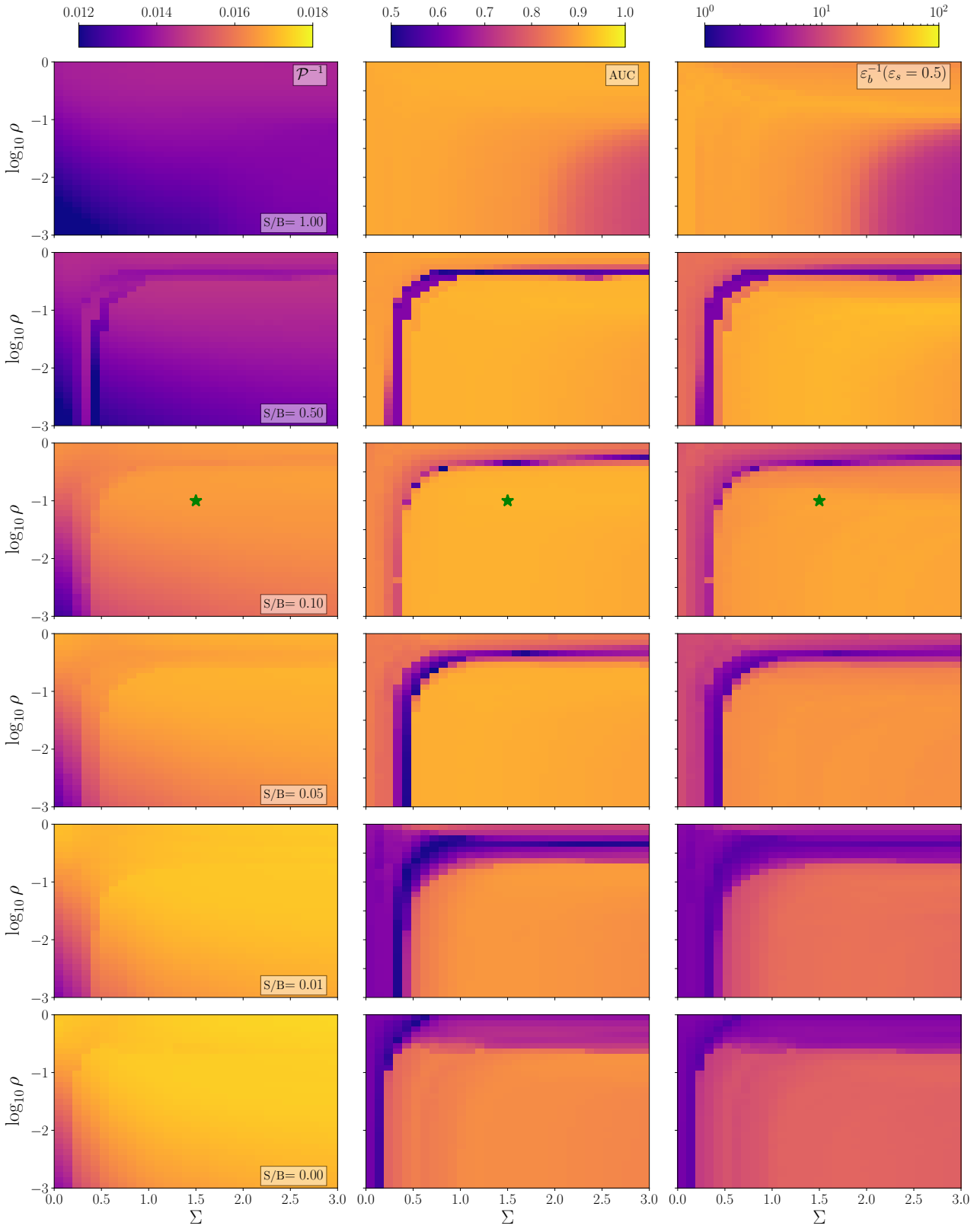



Figure 58: Results of LDA models in the $(\Sigma, \rho)$ parameter-space trained on samples of mixed $t\bar{t}$ and QCD events using mass basis observables $m_0$ and $m_1/m_0$, with different S/B ratios (one per row). Each row contains plots of perplexity, AUC, and inverse mis-tag rate at fixed efficiency (see text for details). The green star indicates the minimum perplexity point for S/B = 0.1. We use this model to plot the extracted theme distributions in Fig. 59.

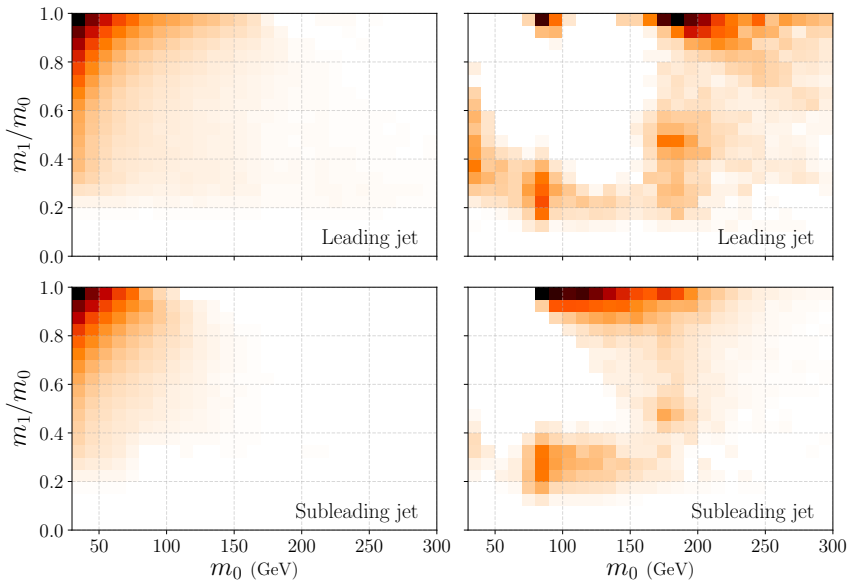



Figure 59: The LDA extracted theme 1 (left) and theme 2 (right) distributions for the leading (upper plots) and subleading (lower plots) jets obtained on a mixed $t\bar{t}$ / QCD sample with S/B = 0.1, where only the $m_0$ and $m_1/m_0$ observables were used. The results for the model that yields the lowest perplexity. See text for details.

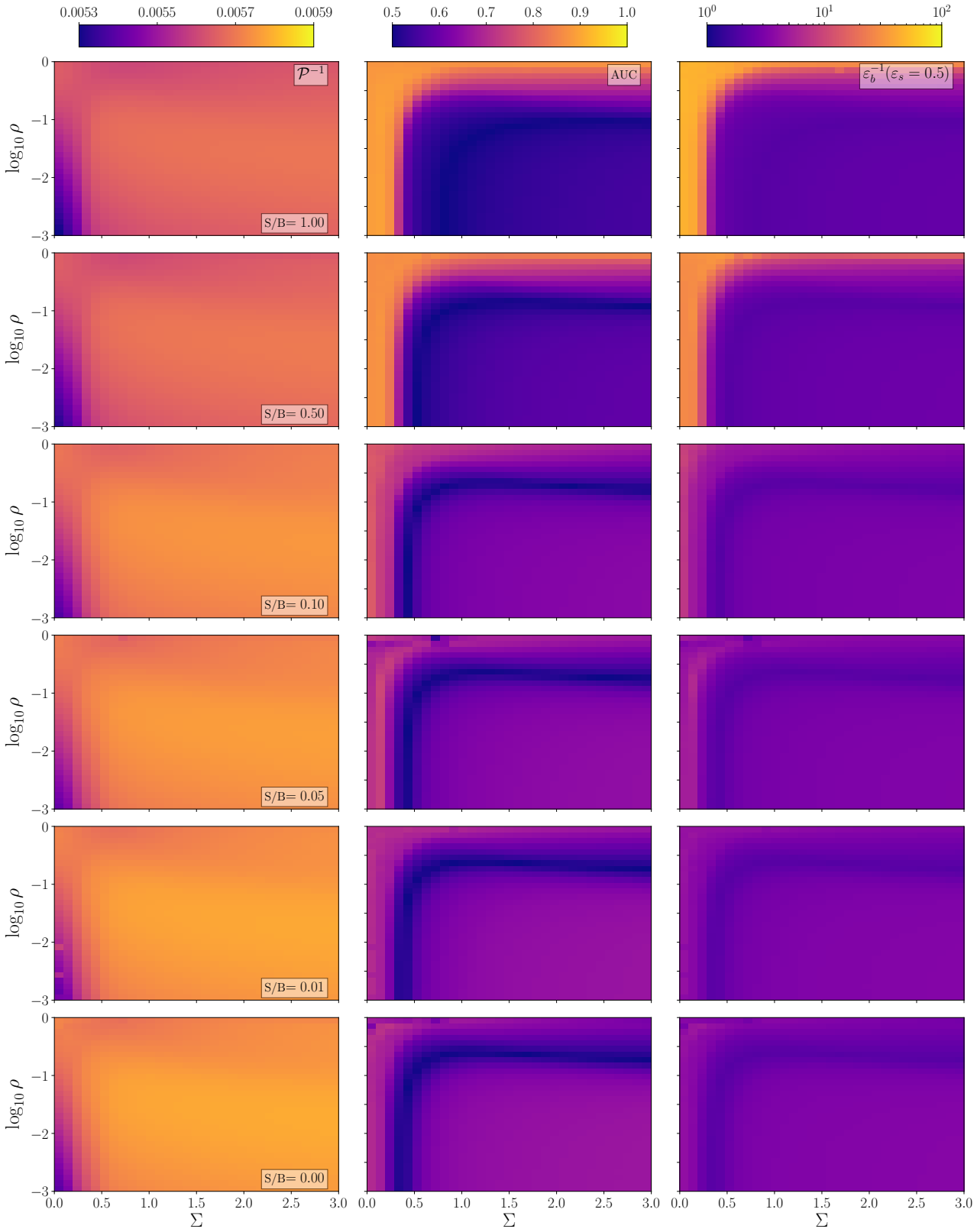



Figure 60: Results of LDA models in the $(\Sigma, \rho)$ parameter-space trained on samples of mixed $t\bar{t}$ and QCD events using Lund basis observables $\log k_T$ and $\log R/\Delta$, with different S/B ratios (one per row). Each row contains plots of perplexity, AUC, and inverse mis-tag rate at fixed efficiency. See text for details.

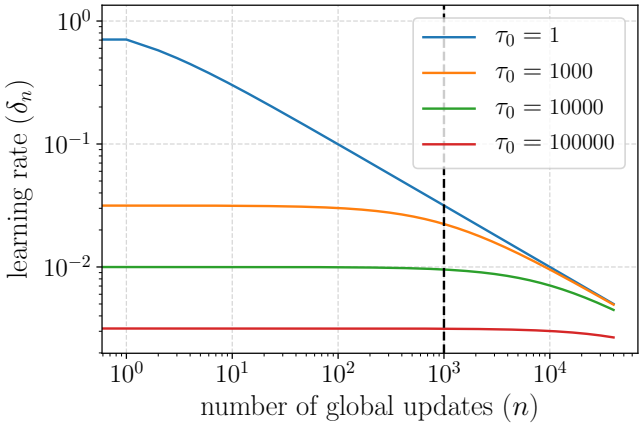



Figure 61: The learning rate as a function of the number of global updates for different offsets, $\tau_0 = 1\text{-}10^5$. The black dashed line indicates the learning rate after 100 passes when we have $10^5$ events in the sample and a chunk size of $10^4$, as in the scans in Sections 6.4.

## 6.5 SYSTEMATICS

In order to use the techniques presented in this Chapter in practice it is important that the VI algorithm produces results which are stable under changes in the random initialisation of the model variables. It is also important to verify that the algorithm parameters chosen for the inference procedure detailed in Subsection 6.1.3 are sensible choices for the datasets studied. The most important algorithm parameters for LDA are the offset $\tau_0$, the chunk size $n_c$, and the number of passes $n_p$. The offset affects the learning rate, both the overall magnitude and as a function of the global updates, see Eq. 105. The chunk size changes how many events are used to optimise the local parameters before an update on the global parameters is performed. Finally, the number of passes must simply be large enough for the algorithm to converge.

### 6.5.1 *Offset*

We show in Fig. 61 how the offset affects the learning rate $\delta_n$, and thus the inference algorithm, as a function of the number of global updates. We see that a larger offset implies a smaller learning rate that is also more constant across the global updates. A smaller offset leads to very large learning rates at earlier global updates, and larger learning rates overall.

To demonstrate the effect that the learning rate and offset have on the results, we have chosen a single parameter point from the scans performed on the $W'$/QCD mixed event sample, with $S/B = 0.025$ and $(\Sigma, \rho) = (1.3, 0.05)$ in the mass basis. We keep all the parameters as they were in the scan except for the offset that is now varied from 1 to $2\times10^5$. For each offset we measure the algorithm performance by calculating the perplexity, the AUC, and the inverse mistag rate. To assess the stability of the algorithm as a function of offset we repeat this for 100 different random seeds, calculating the mean and the upper and lower standard deviations of the resulting distribution. These results are shown in Fig. 62.

A clear effect is that both the perplexity and the performance of these models increase with the size of the offset, degrading heavily at low offsets. This is due to the

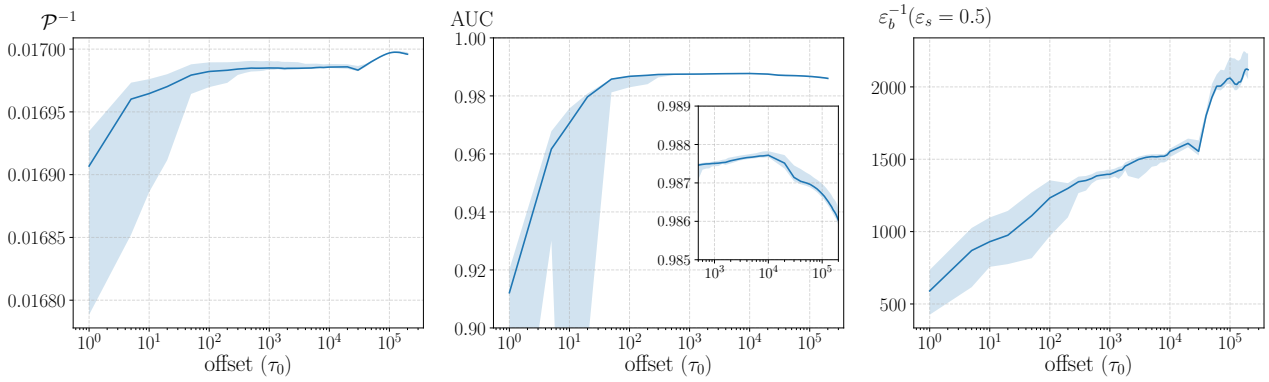



Figure 62: The inverse perplexity (left), AUC (center), and inverse mistag rate (right) as functions of the offset for an event sample with $10^5$ events and a chunk size of $10^4$. The calculations were done for 100 different random seeds, the blue lines show the mean of these and the shaded regions cover the upper and lower standard deviations. Separate upper and lower standard deviations are used to show how the actual variances in the performance statistics are typically skewed heavily towards the negative side.

learning rate being large to sufficiently resolve the maximum in the ELBO. We also see that the random seed induced variance of the results increases considerably at low offsets. This is partially due to the overall size of the learning rate but is also affected by the significantly increased learning rate in the initial global updates, as can be seen in Fig. 61. Because the chunks of data are sampled randomly at the beginning of the analysis, a different random seed means that a different subset of the data will have more influence on the inference, hence the larger variance. Another effect observed is the change in behaviour at very large offset. The AUC and inverse mistag are both good measures of performance for the model so we might expect that an increase in one leads to an increase in the other. However, we see here that this is not the case. At offset $\sim 10^4$ the AUC begins to degrade while the inverse mistag at fixed efficiency of $\varepsilon_s = 0.5$ continues to improve somewhat.

The learning rate also affects the speed of convergence of VI. In the algorithm described in Subsection 6.1.3 we allow the algorithm to run for a fixed number of passes over the data without checking for convergence. However one could easily change this to check explicitly for convergence and end the algorithm early. In Fig. 63 we look at how many passes over the data the algorithm takes to converge, seeing that runs with larger offsets take much longer to converge due to the smaller learning rate.

From these observations we deduce that an offset in the range $10^3$-$10^4$ is the best choice for both the performance and stability of the inference algorithm as applied to our example datasets. Correspondingly, the suitability of a particular offset choice on other datasets can be readily verified by checking for convergence as well as model perplexity dependence on this algorithm parameter.

### 6.5.2 *Chunk size*

In Section 6.4, the chunk size is fixed to $10^4$ while the samples contain almost $10^5$ events, yielding 10 chunks per pass over the sample. Since we are looking for rare signals it is possible that the signal events could be very unevenly distributed throughout these different chunks, resulting in each chunk having significantly different ELBO

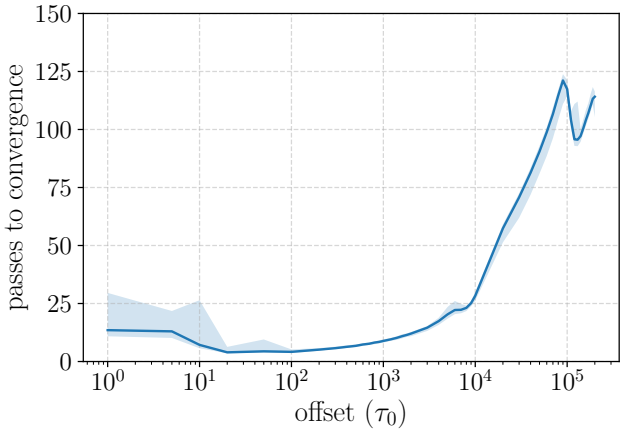



Figure 63: Number of passes needed for convergence as a function of the offset, for an event sample with $10^5$ events and a chunk size of $10^4$. The calculations were done for 100 different random seeds, the blue lines show the mean of these and the shaded regions cover the upper and lower standard deviations. Separate upper and lower standard deviations are used to show how the actual variances in the number of required passes are skewed with respect to the mean.

and perplexity. Therefore, the algorithm would essentially be attempting to optimise one model for these 10 different chunks, and the resulting posterior approximation would fail to accurately describe the true posterior. To test that this is not an issue in the scans, we perform the same offset scan as in Fig. 62 but now for a smaller event sample ($10^4$ events), where the global updates are performed only after seeing the whole dataset as the chunk size is equal to the size of the event sample. We see the results in Fig. 64 and it is clear that while they differ slightly, qualitatively the same behavior is observed in the perplexity and performance at different offsets.

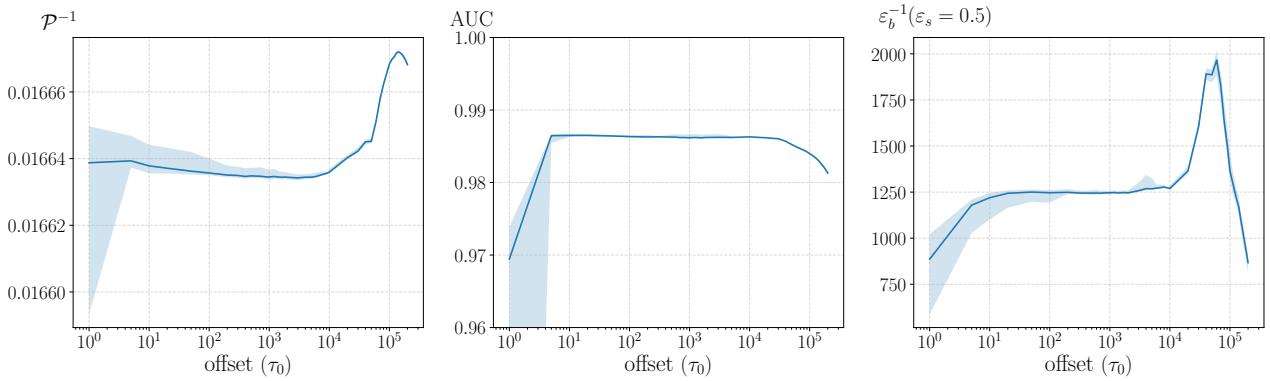



Figure 64: The inverse perplexity (left), AUC (center), and inverse mistag rate (right) as functions of the offset for an event sample with $10^4$ events and a chunk size equal to the size of the event sample. The calculations were done for 100 different random seeds, the blue lines show the mean of these and the shaded regions cover the upper and lower standard deviations. Separate upper and lower standard deviations are used to show how the actual variances in the performance statistics are typically skewed heavily towards the negative side.

To properly study the effect of changing the chunk size we need to find a better way to compare models trained with different chunk sizes. Changing the chunk size significantly affects how much of the data the algorithm analyses before it converges,

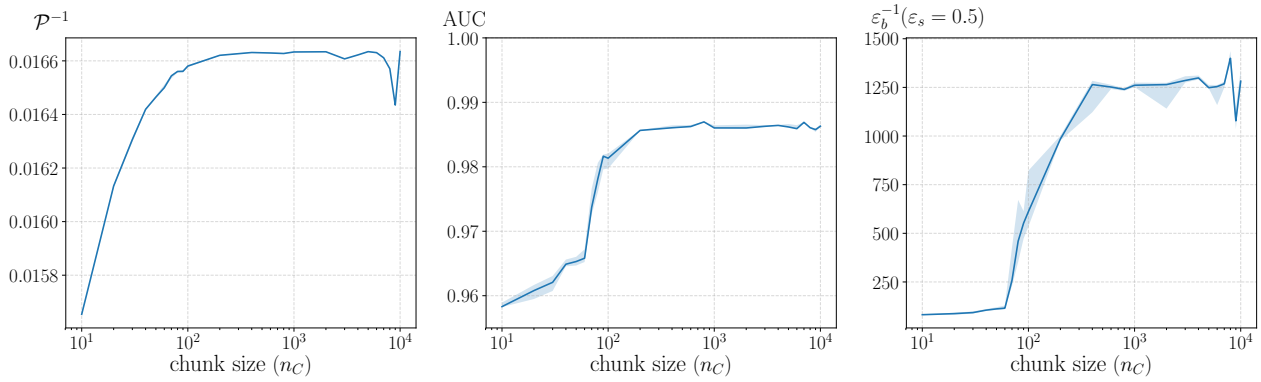



Figure 65: The inverse perplexity (left), AUC (center), and inverse mistag rate (right) as functions of the chunk size for a mixed $W'$/QCD event sample with $10^4$ events and a chunk size equal to the size of the event sample. The offset is also varied such that the learning rate is held constant. The calculations were done for 100 different random seeds, the blue lines show the mean of these and the shaded regions cover the upper and lower standard deviations. Separate upper and lower standard deviations are used to show how the actual variances in the performance statistics are typically skewed heavily towards the negative side.

and we would like to disentangle this effect from the effect due to less data being analysed per global update in the algorithm. To do this we vary the offset simultaneously with the chunk size such that the learning rate at one pass over the data is held constant. The example we use is again the $W'$/QCD mixed sample of a total $10^4$ events, with $S/B = 0.025$ and $(\Sigma, \rho) = (1.3, 0.05)$ in the mass basis. The learning rate is held constant to what it would be if we had an offset of $10^3$ and a chunk size of $10^4$ events. The chunk size is varied from 10 up to $10^4$, meaning that the offset varies from 1 to 1000. The results are shown in Fig. 65, where we clearly see the disadvantages in using very small chunk sizes. When the chunk size reaches $\mathcal{O}(5\%)$ of the size of the event sample the perplexity and performance statistics reach a plateau. As we vary the chunk size we see that the results are not very sensitive to the random seed. This is because the learning rate is kept at a constant (small enough) value by varying the offset accordingly.

From this exploration we take that while choosing the chunk size to be equal to the size of the event sample is certainly a good idea, especially with smaller datasets and rare signals, for the event samples that we have analysed in this Chapter setting the chunk size to only a fraction of total dataset does not significantly impede the quality or robustness of VI while significantly speeding its convergence. In particular, our reasoning in choosing the chunk size to be $10^4$ rather than $10^5$ for our prior scans in Section 6.4 is simply that the algorithm converges 10 times faster.

## 6.6 CONCLUSIONS

In this Chapter we have described a general unsupervised framework capable of learning rare patterns in event data collected at high-energy colliders. We use a Bayesian probabilistic modeling technique called Latent Dirichlet Allocation (LDA), an unsupervised ML approach that was first introduced in the context of BSM collider physics in Ref. [100].

We started by representing individual collider events as sequences of binned exchangeable measurements, and assumed a simplified picture in which the events are generated by sampling these measurements from some underlying joint probability distribution. The assumption of exchangeability of measurements guarantees, through de Finnetti's theorem, that the sequence of measurements in an event are conditionally dependent on a latent variable sampled from a (marginalised over) prior distribution over a latent space. With the help of some basic assumptions on this latent space we arrived at the LDA model, which we focus on throughout the Chapter.

LDA is a mixed-membership model, meaning that under this model the measurements in individual events are assumed to have been sampled from multiple (two, in our case) multinomial distributions, the themes. These themes encode information on the underlying structure in the event data represented in terms of binned measurements. The mixing proportions of themes are sampled from a prior taking the form of a Dirichlet distribution, a parametric family of distributions over the simplex. Mixed membership models have the advantage of describing different events which share features arising from the same underlying physical source. Depending on the Dirichlet prior, the generative model can naturally describe event samples where certain combinations of measurements appear rarely, which is crucial for uncovering rare signals.

Given the LDA model and the event data, we described in detail a stochastic variational inference technique for approximating or learning the underlying themes from which the data is assumed to have been generated. We then described how the extracted themes can be used to construct a classifier to cluster events into two categories, potentially aligned with the background and signal classes. We finally identified a measure of classification performance based solely on the learned themes, the perplexity, which does not require truth labels to compute and can thus be extracted directly from mixed data. In particular, we found that perplexity correlates strongly with the widely used traditional measures of supervised classification performance based on the ROC curves: the AUC and inverse-mistag rate at fixed efficiency.

To demonstrate the power of this technique we considered the analysis of di-jet events at the LHC focusing on two benchmark examples; boosted SM $t\bar{t}$ production and a hypothetical BSM production of $W' \to (\phi \to WW)W$. We described in detail how to pre-process the event data to express each event as a sequence of exchangeable measurements and how the generative model for di-jet events is to be interpreted using LDA. Our choice of jet substructure observables that we used in the analysis is based upon high level observable combinations that have previously been shown to be good for identifying massive resonance decay chains within large radius jets with supervised methods: the traditional mass drop basis (Ref. [233]) and the primary Lund plane basis (Ref. [230]).We note in passing the in principle LDA can be trained on any general combination of high-level observables used in supervised classification that has significant discriminating power, thus in principle allowing to promote supervised classifiers to unsupervised ones, given enough measurement co-occurrences in the data. Through a study of the classification power of these different observables, and of how strong their co-occurrences are in the data, we have identified most promising pairs of observables in each basis for our unsupervised classification approach.

The results for each of the benchmark di-jet examples from this study are presented in Section 6.4. Plotting the perplexity, AUC, and inverse-mistag rate at a fixed signal efficiency as performance indicators, we analysed how well the two-theme LDA models classified events over a large range of values of Dirichlet prior parameters

$(\Sigma, \rho)$. For each benchmark we considered six different samples with varying S/B, ranging from 0.01 to 1.0 for the boosted top-quark example, and from 0.005 to 0.1 for the $W'$ example, adding background only samples for reference. For both benchmark we found that the two bases lead to complementary results, with the extracted themes in each case holding valuable information about the signal and background processes. From the results it is clear that the inference algorithm was able to separate measurement patterns corresponding to the massive resonance decays within the signal jets from patterns corresponding to light QCD emissions present within all jets. This is achieved due to the mixed-membership nature of the generative model, where QCD-like patterns found both in the signal and background jets were identified as having been sampled from the same theme describing QCD-like splittings in the jet substructure.

Finally, in Section 6.5 we studied how the results and performance of the chosen inference technique depend on the tunable parameters of the algorithm, in particular the chunk size and the offset. We demonstrated that the results of the algorithm are in fact stable over a large range of these parameters, and that the algorithm tends to converge within $\lesssim 100$ passes for the example datasets.

Perhaps the most important result of this Chapter is that the AUC and inverse-mistag rate, calculated using truth label information, are strongly correlated over the $(\Sigma, \rho)$ Dirichlet parameter plane with the perplexity, which is calculated without truth label information. This implies that not only can perplexity be used as a practical measure to assess LDA model convergence, but it can also provide guidance when selecting the most viable and robust Dirichlet priors for unsupervised collider analyses and searches. By allowing the algorithm to select optimal $\Sigma$ and $\rho$ parameters we would not need to perform a search for each choice of parameters considered, meaning that there would be no contribution to the trials factor due to these parameters. This is similar to what happens in Ref. [97], where the ATLAS collaboration published an analysis of a weakly supervised di-jet resonance search in which contributions to the trials factor associated to the masses of the final state jets are eliminated by allowing the ML algorithm to define the classifier using the event data alone.

After this detailed study of the LDA method applied to BSM searches in di-jets we can conclude that probabilistic latent variable models offer a unique and promising avenue for future study. The first significant result is that the perplexity, related to the model evidence, can be used as a model-independent measure to select the optimal hyper-parameters for the LDA model, which are connected with the abundances of signal and background-like features in the data. We also presented further understanding of how this method works in terms of feature co-occurrences in the data, and how these affect the performance of the variational inference algorithm. This intuitive understanding illuminates a drawback in the approach, in that we are limited to using a small number of physical observables in the analysis due to the infamous curse of dimensionality, and the associated sparsity of individual measurements. The results in this Chapter provide a theoretical underpinning of the technique and results first presented in Ref. [100], and pave the way for future collider studies using probabilistic latent variable models such as LDA.

# 7

# CONCLUSIONS

In this Ph.D. Thesis we aim to fulfill the goals stated in Chapter 1, exploring in different ways how BSM could be lurking at the LHC. We study different BSM models in Chapters 2, 3 and 4 and different Machine Learning algorithms that reduce Monte Carlo dependency and provide a way towards semi- or unsupervised searches in Chapters 5 and 6.

In Chapter 2 we analyse how an scalar Leptoquark motivated by the B-Anomalies on the experimental side and the Partial Compositeness hypothesis on the theoretical side can be probed by pair production searches at the LHC. We found that existing searches constrain the Leptoquark mass to be higher than $\approx 800$ GeV. We studied how these same searches could constrain the Leptoquark for $\sqrt{s} = 14$ TeV and $\mathcal{L} = 300$ fb$^{-1}$ and we proposed the $t\tau b\nu$ channel as a very interesting target to explore. This channel was explored in Ref. [128].

In Chapter 3 we are motivated by a discarded excess in a second generation Leptoquark pair production analysis to study how a Leptoquark model with a few different characteristics could be hiding in pair production searches and produce the exact type of events that are discarded due to theoretical considerations. We show how a multi-generation Leptoquark with couplings large enough that the single- and non-resonant production modes are relevant could yield an excess in $\mu\nu jj$ without a corresponding excess in $\mu\mu jj$ and without the expected resonance in the mass spectra. We show how flavour constraints can be avoided by considering a multiple Leptoquark scenario and taking advantage that flavour experiments are more sensible to higher Leptoquark masses than ATLAS and CMS.

In Chapter 4 we demonstrate how a top-philic $Z'$ vector boson with flavour-changing couplings could be hiding behind recent multilepton discrepancies between the expected and observed events. In particular, we show how it could mimic both $t\bar{t}W^{\pm}$ and 4-top production in the relevant signal regions and how a broadband search can disentangle possible BSM effects from SM processes. We suggest that the 4-top total leptonic charge imbalance is an useful quantity that could be reported by the experimental collaborations and we propose a kinematic variable that is able to disentangle $tZ'$ from $t\bar{t}W^{\pm}$, MaxMin$(\ell, b)$.

In Chapter 5 we adapt the *Demixer algorithm* to 4-top searches. We find that the nature of the problem at hand makes it challenging to implement the algorithm straightforwardly as in quark/gluon tagging. This is due to the fact that 4-top and its background are physical processes and cannot be defined by the output of the algorithm, as is the case with quark and gluon jets. However, we find that the algorithm is still very useful as it can be used to extract the background distribution and tune the Monte Carlo without the need of a control region.

In Chapter 6 we explore in detail how Latent Dirichlet Allocation (LDA) can be applied to unsupervised searches at the LHC. We focus on two benchmarks which produce di-jet final states, QCD vs $t\bar{t}$ and QCD vs $W'$. We see that the choice of vocabulary is crucial as it must balance discriminating power with co-occurrences in

the corpus. We explore the different hyperparameters of the model. We show how typical values of hyperparameters like chunksize and offset can be chosen and how to explore the landscape of LDA classifiers defined by the weight priors $\alpha$. We find that perplexity is correlated with the two metrics we consider, AUC and inverse mistag at 50 efficiency, and because of this can be an useful metric for unsupervised $\alpha$ without the need to introduce trial factors.

We find that these different approaches to BSM searches at the LHC are complementary. Because there is a vast array of BSM possibilities with a correspondingly vast array of possible final states, semi- and unsupervised techniques could be fundamental to point the way forward. However, the use of these algorithms is not without underlying hypotheses. The application of Machine Learning algorithms to LHC data highlights the importance of data representation. This issue, already present at supervised searches, is key to assessing the sensitivity of the algorithm. Monte Carlo simulations provide a way to validate and choose the appropriate data representation. Moreover, Monte Carlo simulations are still needed to match the obtained distributions with the underlying hard processes that we are interested it. In the language of Chapters 5 and 6, Monte Carlo simulations relate the abstract themes extracted from data to the meaningful themes we are interested in. This is also true when extracting absolute quantities like cross-sections. Even if an anomaly detector yields a set of background and signal events, the cross-sections can only be obtained through an estimation of the acceptances and efficiencies involved. In light of the above, we propose in Chapter 5 how semi- and unsupervised methods could be used to improve Monte Carlo predictions: we indicate that these methods can in principle identify background events and their distributions *in signal regions*, and thus one could tune the Monte Carlo generator using these distributions before contrasting its predictions to the signal candidate events in the same signal region.

Along this Ph.D. Thesis we have performed a set of phenomenological analyses to search for New Physics at the LHC Run II. These studies have been been mainly focused in Leptoquarks, in top-philic BSM contributing to the 4-top final state and in potential $W^{'}$ particles. The investigations have been performed using cut-based fully supervised, and semi-supervised and unsupervised Machine Learning techniques. As we can learn from the obtained results, as well as from the state-of-the-art in the field, luckily there is much to be done. There is a vast array of BSM models that could be hiding in current LHC data and a recent boom in Machine Learning applications to semi-supervised and unsupervised searches, as shown by the LHC Olympics [98]. We are currently exploring other unsupervised algorithms for 4-top physics and improving on LDA with the introduction of topic priors while also working with other latent variable models such as the Hierarchical Dirichlet Process [237].

# 8

# BIBLIOGRAPHY


[1] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. Reading, USA: Addison-Wesley, 1995. ISBN: 978-0-201-50397-5.

[2] Steven Weinberg. *The Quantum theory of fields. Vol. 1: Foundations*. Cambridge University Press, June 2005. ISBN: 978-0-521-67053-1, 978-0-511-25204-4.

[3] Steven Weinberg. *The quantum theory of fields. Vol. 2: Modern applications*. Cambridge University Press, Aug. 2013. ISBN: 978-1-139-63247-8, 978-0-521-67054-8, 978-0-521-55002-4.

[4] Matthew D. Schwartz. *Quantum Field Theory and the Standard Model*. Cambridge University Press, Mar. 2014. ISBN: 978-1-107-03473-0, 978-1-107-03473-0.

[5] P.A. Zyla et al. «Review of Particle Physics.» In: *PTEP* 2020.8 (2020), p. 083C01. DOI: 10.1093/ptep/ptaa104.

[6] Georges Aad et al. «Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC.» In: *Phys. Lett. B* 716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[7] S. Chatrchyan et al. «Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC.» In: *Physics Letters B* 716.1 (Sept. 2012), 30–61. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2012.08.021. URL: http://dx.doi.org/10.1016/j.physletb.2012.08.021.

[8] Albert M Sirunyan et al. «Combined measurements of Higgs boson couplings in proton–proton collisions at $\sqrt{s} = 13$ TeV.» In: *Eur. Phys. J.* C79.5 (2019), p. 421. DOI: 10.1140/epjc/s10052-019-6909-y. arXiv: 1809.10733 [hep-ex].

[9] Ricardo Gonçalo. «Recent highlights in top quark and Higgs boson physics from the LHC.» In: *J. Phys. Conf. Ser.* 1137.1 (2019), p. 012012. DOI: 10.1088/1742-6596/1137/1/012012.

[10] Georges Aad et al. «Measurement of soft-drop jet observables in pp collisions with the ATLAS detector at $\sqrt{s}$ =13 TeV.» In: *Phys. Rev. D* 101.5 (2020), p. 052007. DOI: 10.1103/PhysRevD.101.052007. arXiv: 1912.09837 [hep-ex].

[11] Georges Aad et al. «Test of the universality of $\tau$ and $\mu$ lepton couplings in $W$-boson decays from $t\bar{t}$ events with the ATLAS detector.» In: (July 2020). arXiv: 2007.14040 [hep-ex].

[12] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. «The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations.» In: *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: arXiv:1405.0301 [hep-ph].

[13] Pierre Artoisenet, Rikkert Frederix, Olivier Mattelaer, and Robbert Rietkerk. «Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations.» In: *JHEP* 03 (2013), p. 015. DOI: 10.1007/JHEP03(2013)015. arXiv: 1212.3460 [hep-ph].

[14] Adam Alloul, Neil D. Christensen, Céline Degrande, Claude Duhr, and Benjamin Fuks. «FeynRules 2.0 - A complete toolbox for tree-level phenomenology.» In: *Comput. Phys. Commun.* 185 (2014), p. 2250. DOI: 10.1016/j.cpc.2014.04.012. arXiv: arXiv:1310.1921 [hep-ph].

[15] Johan Alwall et al. «A Standard format for Les Houches event files.» In: *Comput. Phys. Commun.* 176 (2007), pp. 300–304. DOI: 10.1016/j.cpc.2006.11.010. arXiv: hep-ph/0609017.

[16] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. «An Introduction to PYTHIA 8.2.» In: *Comput. Phys. Commun.* 191 (2015), p. 159. DOI: 10.1016/j.cpc.2015.01.024. arXiv: arXiv:1410.3012 [hep-ph].

[17] Matt Dobbs and Jorgen Beck Hansen. «The HepMC C++ Monte Carlo event record for High Energy Physics.» In: *Comput. Phys. Commun.* 134 (2001), pp. 41–46. DOI: 10.1016/S0010-4655(00)00189-2.

[18] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi. «DELPHES 3, A modular framework for fast simulation of a generic collider experiment.» In: *JHEP* 02 (2014), p. 057. DOI: 10.1007/JHEP02(2014)057. arXiv: arXiv:1307.6346 [hep-ex].

[19] R. Brun and F. Rademakers. «ROOT: An object oriented data analysis framework.» In: *Nucl. Instrum. Meth. A* 389 (1997). Ed. by M. Werlen and D. Perret-Gallix, p. 81. DOI: 10.1016/S0168-9002(97)00048-X.

[20] Enrico Bothmann et al. «Event Generation with Sherpa 2.2.» In: *SciPost Phys.* 7.3 (2019), p. 034. DOI: 10.21468/SciPostPhys.7.3.034. arXiv: 1905.09127 [hep-ph].

[21] M. Bahr et al. «Herwig++ Physics and Manual.» In: *Eur. Phys. J. C* 58 (2008), pp. 639–707. DOI: 10.1140/epjc/s10052-008-0798-9. arXiv: 0803.0883 [hep-ph].

[22] Johannes Bellm et al. «Herwig 7.0/Herwig++ 3.0 release note.» In: *Eur. Phys. J. C* 76.4 (2016), p. 196. DOI: 10.1140/epjc/s10052-016-4018-8. arXiv: 1512.01178 [hep-ph].

[23] S. Agostinelli et al. «GEANT4–a simulation toolkit.» In: *Nucl. Instrum. Meth. A* 506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8.

[24] Rikkert Frederix, Davide Pagani, and Marco Zaro. «Large NLO corrections in $t\bar{t}W^{\pm}$ and $t\bar{t}t\bar{t}$ hadroproduction from supposedly subleading EW contributions.» In: *JHEP* 02 (2018), p. 031. DOI: 10.1007/JHEP02(2018)031. arXiv: arXiv:1711.02116 [hep-ph].

[25] ATLAS Collaboration. «Evidence for $t\bar{t}t\bar{t}$ production in the multilepton final state in proton-proton collisions at $\sqrt{s}$=13 TeV with the ATLAS detector.» In: (July 2020). arXiv: arXiv:2007.14858 [hep-ex].

[26] CMS Collaboration. «Search for production of four top quarks in final states with same-sign or multiple leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV.» In: *Eur. Phys. J. C* 80.2 (2020), p. 75. DOI: 10.1140/epjc/s10052-019-7593-7. arXiv: arXiv:1908.06463 [hep-ex].

[27] Ezequiel Alvarez, Aurelio Juste, and Rosa María Sandá Seoane. «Four-top as probe of light top-philic New Physics.» In: (2019). arXiv: 1910.09581 [hep-ph].

[28] Ezequiel Alvarez, Darius A. Faroughy, Jernej F. Kamenik, Roberto Morales, and Alejandro Szynkman. «Four Tops for LHC.» In: *Nucl. Phys.* B915 (2017), pp. 19–43. DOI: 10.1016/j.nuclphysb.2016.11.024. arXiv: 1611.05032 [hep-ph].

[29] Christoph Englert, Gian F. Giudice, Admir Greljo, and Matthew McCullough. «The $\hat{H}$-parameter: an oblique Higgs view.» In: *Journal of High Energy Physics* 2019.9 (Sept. 2019). ISSN: 1029-8479. DOI: 10.1007/jhep09(2019)041. URL: http://dx.doi.org/10.1007/JHEP09(2019)041.

[30] P. Azzi et al. «Standard Model Physics at the HL-LHC and HE-LHC.» In: (2019). arXiv: 1902.04070 [hep-ph].

[31] Samuel Calvet, Benjamin Fuks, Philippe Gris, and Loïc Valéry. «Searching for sgluons in multitop events at a center-of-mass energy of 8 TeV.» In: *Journal of High Energy Physics* 2013.4 (Apr. 2013). ISSN: 1029-8479. DOI: 10.1007/jhep04(2013)043. URL: http://dx.doi.org/10.1007/JHEP04(2013)043.

[32] Luc Darmé, Benjamin Fuks, and Mark Goodsell. «Cornering sgluons with four-top-quark events.» In: *Physics Letters B* 784 (Sept. 2018), 223–228. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2018.08.001. URL: http://dx.doi.org/10.1016/j.physletb.2018.08.001.

[33] Cen Zhang. «Constraining qqtt operators from four-top production: a case for enhanced EFT sensitivity.» In: *Chinese Physics C* 42.2 (Feb. 2018), p. 023104. ISSN: 1674-1137. DOI: 10.1088/1674-1137/42/2/023104. URL: http://dx.doi.org/10.1088/1674-1137/42/2/023104.

[34] Otilia Ducu, Lucien Heurtier, and Julien Maurer. «LHC signatures of a Z' mediator between dark matter and the SU(3) sector.» In: *JHEP* 03 (2016), p. 006. DOI: 10.1007/JHEP03(2016)006. arXiv: 1509.05615 [hep-ph].

[35] I. Doršner, S. Fajfer, A. Greljo, J. F. Kamenik, and N. Košnik. «Physics of leptoquarks in precision experiments and at particle colliders.» In: *Phys. Rept.* 641 (2016), p. 1. DOI: 10.1016/j.physrep.2016.06.001. arXiv: arXiv:1603.04993 [hep-ph].

[36] Mario Antonelli et al. «Flavor Physics in the Quark Sector.» In: *Phys. Rept.* 494 (2010), pp. 197–414. DOI: 10.1016/j.physrep.2010.05.003. arXiv: 0907.5386 [hep-ph].

[37] LHCb Collaboration. «Test of lepton universality using $B^+ \to K^+\ell^+\ell^-$ decays.» In: *Phys. Rev. Lett.* 113 (2014), p. 151601. DOI: 10.1103/PhysRevLett.113.151601. arXiv: arXiv:1406.6482 [hep-ex].

[38] LHCb Collaboration. «Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays.» In: *JHEP* 08 (2017), p. 055. DOI: 10.1007/JHEP08(2017)055. arXiv: arXiv:1705.05802 [hep-ex].

[39] Gudrun Hiller and Ivan Nisandzic. «$R_K$ and $R_{K^*}$ beyond the standard model.» In: *Phys. Rev. D* 96.3 (2017), p. 035003. DOI: 10.1103/PhysRevD.96.035003. arXiv: arXiv:1704.05444 [hep-ph].

[40] Bernat Capdevila, Andreas Crivellin, Sébastien Descotes-Genon, Joaquim Matias, and Javier Virto. «Patterns of New Physics in $b \to s\ell^+\ell^-$ transitions in the light of recent data.» In: *JHEP* 01 (2018), p. 093. DOI: 10.1007/JHEP01(2018)093. arXiv: 1704.05340 [hep-ph].

[41] LHCb Collaboration. «Measurement of the ratio of branching fractions $\mathcal{B}(\bar{B}^0 \to D^{*+}\tau^-\bar{\nu}_\tau)/\mathcal{B}(\bar{B}^0 \to D^{*+}\mu^-\bar{\nu}_\mu)$.» In: *Phys. Rev. Lett.* 115.11 (2015). [Erratum: Phys. Rev. Lett. 115 (2015) 159901], p. 111803. DOI: 10.1103/PhysRevLett.115.159901, 10.1103/PhysRevLett.115.111803. arXiv: arXiv:1506.08614 [hep-ex].

[42] Belle Collaboration. «Measurement of the branching ratio of $\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau$ relative to $\bar{B} \to D^{(*)}\ell^-\bar{\nu}_\ell$ decays with hadronic tagging at Belle.» In: *Phys. Rev. D* 92.7 (2015), p. 072014. DOI: 10.1103/PhysRevD.92.072014. arXiv: arXiv:1507.03233 [hep-ex].

[43] Marat Freytsis, Zoltan Ligeti, and Joshua T. Ruderman. «Flavor models for $\bar{B} \to D^{(*)}\tau\bar{\nu}$.» In: *Phys. Rev. D* 92.5 (2015), p. 054018. DOI: 10.1103/PhysRevD.92.054018. arXiv: 1506.08896 [hep-ph].

[44] Y. Amhis et al. «Averages of b-hadron, c-hadron, and $\tau$-lepton properties as of summer 2016.» In: *Eur. Phys. J. C* 77.12 (2017), p. 895. DOI: 10.1140/epjc/s10052-017-5058-4. arXiv: 1612.07233 [hep-ex].

[45] Svjetlana Fajfer. «Scalar (or vector) leptoquarks in B meson anomalies.» In: *Nucl. Part. Phys. Proc.* 285 (2017), p. 81. DOI: 10.1016/j.nuclphysbps.2017.03.015.

[46] David Marzocca. «Addressing the B-physics anomalies in a fundamental Composite Higgs Model.» In: *JHEP* 07 (2018), p. 121. DOI: 10.1007/JHEP07(2018)121. arXiv: arXiv:1803.10972 [hep-ph].

[47] A. Angelescu, Damir Bečirević, D. A. Faroughy, and O. Sumensari. «Closing the window on single leptoquark solutions to the B-physics anomalies.» In: (2018). arXiv: 1808.08179 [hep-ph].

[48] Damir Bečirević, Ilja Doršner, Svjetlana Fajfer, Nejc Košnik, Darius A. Faroughy, and Olcyr Sumensari. «Scalar leptoquarks from grand unified theories to accommodate the B-physics anomalies.» In: *Phys. Rev. D* 98.5 (2018), p. 055003. DOI: 10.1103/PhysRevD.98.055003. arXiv: 1806.05689 [hep-ph].

[49] Andreas Crivellin, Dario Muller, and Toshihiko Ota. «Simultaneous explanation of $R(D^{(*)})$ and $b \to s\mu^+\mu^-$: the last scalar leptoquarks standing.» In: *JHEP* 09 (2017), p. 040. DOI: 10.1007/JHEP09(2017)040. arXiv: arXiv:1703.09226 [hep-ph].

[50] Ilja Doršner, Svjetlana Fajfer, Darius A. Faroughy, and Nejc Košnik. «The role of the $S_3$ GUT leptoquark in flavor universality and collider searches.» In: *JHEP* 10 (2017), p. 188. DOI: 10.1007/JHEP10(2017)188. arXiv: arXiv:1706.07779 [hep-ph].

[51] Dario Buttazzo, Admir Greljo, Gino Isidori, and David Marzocca. «B-physics anomalies: a guide to combined explanations.» In: *JHEP* 11 (2017), p. 044. DOI: 10.1007/JHEP11(2017)044. arXiv: arXiv:1706.07808 [hep-ph].

[52] Ben Gripaios, Marco Nardecchia, and S. A. Renner. «Composite leptoquarks and anomalies in B-meson decays.» In: *JHEP* 05 (2015), p. 006. DOI: 10.1007/JHEP05(2015)006. arXiv: arXiv:1412.1791 [hep-ph].

[53] Martin Bauer and Matthias Neubert. «Minimal Leptoquark Explanation for the $R_{D^{(*)}}$, $R_K$, and $(g-2)_g$ Anomalies.» In: *Phys. Rev. Lett.* 116.14 (2016), p. 141802. DOI: 10.1103/PhysRevLett.116.141802. arXiv: 1511.01900 [hep-ph].

[54] Roel Aaij et al. «Test of lepton universality in beauty-quark decays.» In: (Mar. 2021). arXiv: 2103.11769 [hep-ex].

[55] S. Hirose et al. «Measurement of the $\tau$ lepton polarization and $R(D^*)$ in the decay $\bar{B} \to D^* \tau^- \bar{\nu}_\tau$ with one-prong hadronic $\tau$ decays at Belle.» In: *Phys. Rev. D* 97.1 (2018), p. 012004. DOI: 10.1103/PhysRevD.97.012004. arXiv: 1709.00129 [hep-ex].

[56] R. Aaij et al. «Test of Lepton Flavor Universality by the measurement of the $B^0 \to D^{*-} \tau^+ \nu_\tau$ branching fraction using three-prong $\tau$ decays.» In: *Phys. Rev. D* 97.7 (2018), p. 072013. DOI: 10.1103/PhysRevD.97.072013. arXiv: 1711.02505 [hep-ex].

[57] G. Caria et al. «Measurement of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ with a semileptonic tagging method.» In: *Phys. Rev. Lett.* 124.16 (2020), p. 161803. DOI: 10.1103/PhysRevLett.124.161803. arXiv: 1910.05864 [hep-ex].

[58] Bastian Diaz, Martin Schmaltz, and Yi-Ming Zhong. «The leptoquark Hunter's guide: Pair production.» In: *JHEP* 10 (2017), p. 097. DOI: 10.1007/JHEP10(2017)097. arXiv: 1706.05033 [hep-ph].

[59] Nirmal Raj. «Anticipating nonresonant new physics in dilepton angular spectra at the LHC.» In: *Phys. Rev. D* 95.1 (2017), p. 015011. DOI: 10.1103/PhysRevD.95.015011. arXiv: 1610.03795 [hep-ph].

[60] Martin Schmaltz and Yi-Ming Zhong. «The leptoquark Hunter's guide: large coupling.» In: *JHEP* 01 (2019), p. 132. DOI: 10.1007/JHEP01(2019)132. arXiv: 1810.10017 [hep-ph].

[61] Jens Erler, Paul Langacker, Shoaib Munir, and Eduardo Rojas. «Improved Constraints on Z-prime Bosons from Electroweak Precision Data.» In: *JHEP* 08 (2009), p. 017. DOI: 10.1088/1126-6708/2009/08/017. arXiv: arXiv:0906.2435 [hep-ph].

[62] Paul Langacker and Michael Plumacher. «Flavor changing effects in theories with a heavy Z′ boson with family nonuniversal couplings.» In: *Phys. Rev. D* 62 (2000), p. 013006. DOI: 10.1103/PhysRevD.62.013006. arXiv: hep-ph/0001204.

[63] Andrzej J. Buras, Fulvia De Fazio, and Jennifer Girrbach. «The Anatomy of Z′ and Z with Flavour Changing Neutral Currents in the Flavour Precision Era.» In: *JHEP* 02 (2013), p. 116. DOI: 10.1007/JHEP02(2013)116. arXiv: 1211.1896 [hep-ph].

[64] Peter Cox, Anibal D. Medina, Tirtha Sankar Ray, and Andrew Spray. «Novel collider and dark matter phenomenology of a top-philic Z′.» In: *JHEP* 06 (2016), p. 110. DOI: 10.1007/JHEP06(2016)110. arXiv: arXiv:1512.00471 [hep-ph].

[65] Johan Alwall, Claude Duhr, Benjamin Fuks, Olivier Mattelaer, Deniz Gizem Öztürk, and Chia-Hsien Shen. «Computing decay rates for new physics theories with FeynRules and MadGraph5_aMC@NLO.» In: *Comput. Phys. Commun.* 197 (2015), pp. 312–323. DOI: 10.1016/j.cpc.2015.08.031. arXiv: 1402.1178 [hep-ph].

[66] J.I. Aranda, D. Espinosa-Gómez, J. Montaño, F. Ramírez-Zavaleta, and E.S. Tututi. «Rare decays of the top quark mediated by Z′ gauge bosons and flavor violation.» In: *Mod. Phys. Lett. A* 35.18 (2020), p. 2050153. DOI: 10.1142/S0217732320501539. arXiv: arXiv:2005.02334 [hep-ph].

[67] ATLAS Collaboration. «Search for flavour-changing neutral currents in processes with one top quark and a photon using 81 $fb^{-1}$ of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS experiment.» In: *Phys. Lett. B* 800 (2020), p. 135082. DOI: 10.1016/j.physletb.2019.135082. arXiv: arXiv:1908.08461 [hep-ex].

[68] ATLAS Collaboration. «Search for FCNC single top-quark production at $\sqrt{s} =$ 7 TeV with the ATLAS detector.» In: *Phys. Lett. B* 712 (2012), p. 351. DOI: 10.1016/j.physletb.2012.05.022. arXiv: arXiv:1203.0529 [hep-ex].

[69] CMS Collaboration. «Search for Anomalous Single Top Quark Production in Association with a Photon in pp Collisions at $\sqrt{s} = 8$ TeV.» In: *JHEP* 04 (2016), p. 035. DOI: 10.1007/JHEP04(2016)035. arXiv: arXiv:1511.03951 [hep-ex].

[70] CMS Collaboration. «Search for anomalous Wtb couplings and flavour-changing neutral currents in t-channel single top quark production in pp collisions at $\sqrt{s} = 7$ and 8 TeV.» In: *JHEP* 02 (2017), p. 028. DOI: 10.1007/JHEP02(2017)028. arXiv: arXiv:1610.03545 [hep-ex].

[71] Christopher M Bishop. *Pattern recognition and machine learning*. Information science and statistics. Softcover published in 2016. New York, NY: Springer, 2006. URL: https://cds.cern.ch/record/998831.

[72] Eric M. Metodiev, Benjamin Nachman, and Jesse Thaler. «Classification without labels: Learning from mixed samples in high energy physics.» In: *JHEP* 10 (2017), p. 174. DOI: 10.1007/JHEP10(2017)174. arXiv: 1708.02949 [hep-ph].

[73] Stefan von Buddenbrock, Olivier Mattelaer, and Michael Spannowsky. «Towards a generic implementation of matrix-element maximisation as a classifier in particle physics.» In: (2019). arXiv: 1908.05286 [hep-ph].

[74] Jack H. Collins, Kiel Howe, and Benjamin Nachman. «Extending the search for new resonances with machine learning.» In: *Phys. Rev.* D99.1 (2019), p. 014038. DOI: 10.1103/PhysRevD.99.014038. arXiv: 1902.02634 [hep-ph].

[75] Theo Heimel, Gregor Kasieczka, Tilman Plehn, and Jennifer M. Thompson. «QCD or What?» In: *SciPost Phys.* 6 (2019), p. 030. DOI: 10.21468/SciPostPhys.6.3.030. arXiv: 1808.08979 [hep-ph].

[76] Benjamin Nachman and David Shih. «Anomaly Detection with Density Estimation.» In: *Phys. Rev. D* 101 (2020), p. 075042. DOI: 10.1103/PhysRevD.101.075042. arXiv: 2001.04990 [hep-ph].

[77] Anders Andreassen, Benjamin Nachman, and David Shih. «Simulation Assisted Likelihood-free Anomaly Detection.» In: *Phys. Rev. D* 101.9 (2020), p. 095004. DOI: 10.1103/PhysRevD.101.095004. arXiv: 2001.05001 [hep-ph].

[78] Gregor Kasieczka, Nicholas Kiefer, Tilman Plehn, and Jennifer M. Thompson. «Quark-Gluon Tagging: Machine Learning vs Detector.» In: *SciPost Phys.* 6.6 (2019), p. 069. DOI: 10.21468/SciPostPhys.6.6.069. arXiv: 1812.09223 [hep-ph].

[79] Andrea De Simone and Thomas Jacques. «Guiding New Physics Searches with Unsupervised Learning.» In: *Eur. Phys. J.* C79.4 (2019), p. 289. DOI: 10.1140/epjc/s10052-019-6787-3. arXiv: 1807.06038 [hep-ph].

[80] Andrew Blance, Michael Spannowsky, and Philip Waite. «Adversarially-trained autoencoders for robust unsupervised new physics searches.» In: *JHEP* 10 (2019), p. 047. DOI: 10.1007/JHEP10(2019)047. arXiv: 1905.10384 [hep-ph].

[81] Jack H. Collins, Kiel Howe, and Benjamin Nachman. «Anomaly Detection for Resonant New Physics with Machine Learning.» In: *Phys. Rev. Lett.* 121.24 (2018), p. 241803. DOI: 10.1103/PhysRevLett.121.241803. arXiv: 1805.02664 [hep-ph].

[82] Eric M. Metodiev and Jesse Thaler. «Jet Topics: Disentangling Quarks and Gluons at Colliders.» In: *Phys. Rev. Lett.* 120.24 (2018), p. 241602. DOI: 10.1103/PhysRevLett.120.241602. arXiv: 1802.00008 [hep-ph].

[83] Patrick T. Komiske, Eric M. Metodiev, and Jesse Thaler. «An operational definition of quark and gluon jets.» In: *JHEP* 11 (2018), p. 059. DOI: 10.1007/JHEP11(2018)059. arXiv: 1809.01140 [hep-ph].

[84] Patrick T. Komiske, Eric M. Metodiev, Benjamin Nachman, and Matthew D. Schwartz. «Learning to classify from impure samples with high-dimensional data.» In: *Phys. Rev.* D98.1 (2018), p. 011502. DOI: 10.1103/PhysRevD.98.011502. arXiv: 1801.10158 [hep-ph].

[85] Lucio Mwinmaarong Dery, Benjamin Nachman, Francesco Rubbo, and Ariel Schwartzman. «Weakly Supervised Classification in High Energy Physics.» In: *JHEP* 05 (2017), p. 145. DOI: 10.1007/JHEP05(2017)145. arXiv: 1702.00414 [hep-ph].

[86] Timothy Cohen, Marat Freytsis, and Bryan Ostdiek. «(Machine) Learning to Do More with Less.» In: *JHEP* 02 (2018), p. 034. DOI: 10.1007/JHEP02(2018)034. arXiv: 1706.09451 [hep-ph].

[87] Kaustubh Agashe, Jack H. Collins, Peizhi Du, Sungwoo Hong, Doojin Kim, and Rashmish K. Mishra. «Detecting a Boosted Diboson Resonance.» In: *JHEP* 11 (2018), p. 027. DOI: 10.1007/JHEP11(2018)027. arXiv: 1809.07334 [hep-ph].

[88] J. A. Aguilar-Saavedra, Jack H. Collins, and Rashmish K. Mishra. «A generic anti-QCD jet tagger.» In: *JHEP* 11 (2017), p. 163. DOI: 10.1007/JHEP11(2017)163. arXiv: 1709.01087 [hep-ph].

[89] Jan Hajer, Ying-Ying Li, Tao Liu, and He Wang. «Novelty Detection Meets Collider Physics.» In: *Phys. Rev. D* 101.7 (2020), p. 076015. DOI: 10.1103/PhysRevD.101.076015. arXiv: 1807.10261 [hep-ph].

[90] Marco Farina, Yuichiro Nakai, and David Shih. «Searching for New Physics with Deep Autoencoders.» In: *Phys. Rev. D* 101.7 (2020), p. 075021. DOI: 10.1103/PhysRevD.101.075021. arXiv: 1808.08992 [hep-ph].

[91] Olmo Cerri, Thong Q. Nguyen, Maurizio Pierini, Maria Spiropulu, and Jean-Roch Vlimant. «Variational Autoencoders for New Physics Mining at the Large Hadron Collider.» In: *JHEP* 05 (2019), p. 036. DOI: 10.1007/JHEP05(2019)036. arXiv: 1811.10276 [hep-ex].

[92] Raffaele Tito D'Agnolo and Andrea Wulzer. «Learning New Physics from a Machine.» In: *Phys. Rev. D* 99.1 (2019), p. 015014. DOI: 10.1103/PhysRevD.99.015014. arXiv: 1806.02350 [hep-ph].

[93] Raffaele Tito D'Agnolo, Gaia Grosso, Maurizio Pierini, Andrea Wulzer, and Marco Zanetti. «Learning Multivariate New Physics.» In: (Dec. 2019). arXiv: 1912.12155 [hep-ph].

[94] Oz Amram and Cristina Mantilla Suarez. «Tag N' Train: A Technique to Train Improved Classifiers on Unlabeled Data.» In: (Feb. 2020). arXiv: 2002.12376 [hep-ph].

[95] M. Crispim Romao, N.F. Castro, J.G. Milhano, R. Pedro, and T. Vale. «Use of a Generalized Energy Mover's Distance in the Search for Rare Phenomena at Colliders.» In: (Apr. 2020). arXiv: 2004.09360 [hep-ph].

[96] Oliver Knapp, Guenther Dissertori, Olmo Cerri, Thong Q. Nguyen, Jean-Roch Vlimant, and Maurizio Pierini. «Adversarially Learned Anomaly Detection on CMS Open Data: re-discovering the top quark.» In: (May 2020). arXiv: 2005.01598 [hep-ex].

[97] Georges Aad et al. «Dijet resonance search with weak supervision using $\sqrt{s} =$ 13 TeV pp collisions in the ATLAS detector.» In: (May 2020). arXiv: 2005.02983 [hep-ex].

[98] Gregor Kasieczka et al. «The LHC Olympics 2020: A Community Challenge for Anomaly Detection in High Energy Physics.» In: (Jan. 2021). arXiv: 2101.08320 [hep-ph].

[99] S. Deerwester, S. Dumais, T. Landauer, G. Furnas, and R. Harshman. «Indexing by Latent Semantic Analysis.» In: *J. Am. Soc. Inf. Sci.* 41 (1990), pp. 391–407.

[100] Barry M. Dillon, Darius A. Faroughy, and Jernej F. Kamenik. «Uncovering latent jet substructure.» In: *Phys. Rev. D* 100.5 (2019), p. 056002. DOI: 10.1103/PhysRevD.100.056002. arXiv: 1904.04200 [hep-ph].

[101] Ben Gripaios. «Composite Leptoquarks at the LHC.» In: *JHEP* 02 (2010), p. 045. DOI: 10.1007/JHEP02(2010)045. arXiv: arXiv:0910.1789 [hep-ph].

[102] Ezequiel Alvarez, Leandro Da Rold, Aurelio Juste, Manuel Szewc, and Tamara Vazquez Schroeder. «A composite pNGB leptoquark at the LHC.» In: *JHEP* 12 (2018), p. 027. DOI: 10.1007/JHEP12(2018)027. arXiv: 1808.02063 [hep-ph].

[103] Leandro Da Rold and Federico Lamagna. «Composite Higgs and leptoquarks from a simple group.» In: *JHEP* 03 (2019), p. 135. DOI: 10.1007/JHEP03(2019)135. arXiv: 1812.08678 [hep-ph].

[104] Tony Gherghetta and Alex Pomarol. «Bulk fields and supersymmetry in a slice of AdS.» In: *Nucl. Phys. B* 586 (2000), p. 141. DOI: 10.1016/S0550-3213(00)00392-8. arXiv: arXiv:hep-ph/0003129 [hep-ph].

[105] Csaba Csaki, Adam Falkowski, and Andreas Weiler. «The Flavor of the Composite Pseudo-Goldstone Higgs.» In: *JHEP* 09 (2008), p. 008. DOI: 10.1088/1126-6708/2008/09/008. arXiv: arXiv:0804.1954 [hep-ph].

[106] G. F. Giudice, C. Grojean, A. Pomarol, and R. Rattazzi. «The Strongly-Interacting Light Higgs.» In: *JHEP* 06 (2007), p. 045. DOI: 10.1088/1126-6708/2007/06/045. arXiv: arXiv:hep-ph/0703164 [hep-ph].

[107] David B. Kaplan. «Flavor at SSC energies: A New mechanism for dynamically generated fermion masses.» In: *Nucl. Phys. B* 365 (1991), p. 259. DOI: 10.1016/S0550-3213(05)80021-5.

[108] Boaz Keren-Zur, Paolo Lodone, Marco Nardecchia, Duccio Pappadopulo, Riccardo Rattazzi, and Luca Vecchi. «On Partial Compositeness and the CP asymmetry in charm decays.» In: *Nucl. Phys. B* 867 (2013), p. 394. DOI: 10.1016/j.nuclphysb.2012.10.012. arXiv: arXiv:1205.5803 [hep-ph].

[109] Martin Bauer, Raoul Malm, and Matthias Neubert. «A Solution to the Flavor Problem of Warped Extra-Dimension Models.» In: *Phys. Rev. Lett.* 108 (2012), p. 081603. DOI: 10.1103/PhysRevLett.108.081603. arXiv: arXiv:1110.0471 [hep-ph].

[110] Leandro Da Rold and Iván A. Davidovich. «A symmetry for $\epsilon_K$.» In: *JHEP* 10 (2017), p. 135. DOI: 10.1007/JHEP10(2017)135. arXiv: arXiv:1704.08704 [hep-ph].

[111] Leandro Da Rold. «Anarchy with linear and bilinear interactions.» In: *JHEP* 10 (2017), p. 120. DOI: 10.1007/JHEP10(2017)120. arXiv: arXiv:1708.08515 [hep-ph].

[112] Giuliano Panico and Andrea Wulzer. «The Composite Nambu-Goldstone Higgs.» In: *Lect. Notes Phys.* 913 (2016), pp.1–316. DOI: 10.1007/978-3-319-22617-0. arXiv: arXiv:1506.01961 [hep-ph].

[113] Albert M Sirunyan et al. «Search for third-generation scalar leptoquarks and heavy right-handed neutrinos in final states with two tau leptons and two jets in proton-proton collisions at $\sqrt{s} = 13$ TeV.» In: *JHEP* 07 (2017), p. 121. DOI: 10.1007/JHEP07(2017)121. arXiv: arXiv:1703.03995 [hep-ex].

[114] ATLAS Collaboration. «Search for B-L R-parity-violating top squarks in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS experiment.» In: *Phys. Rev. D* 97 (2018), p. 032003. DOI: 10.1103/PhysRevD.97.032003. arXiv: arXiv:1710.05544 [hep-ex].

[115] CMS Collaboration. «Search for new phenomena with the $M_{T2}$ variable in the all-hadronic final state produced in proton-proton collisions at $\sqrt{s} = 13$ TeV.» In: *Eur. Phys. J. C* 77.10 (2017), p. 710. DOI: 10.1140/epjc/s10052-017-5267-x. arXiv: arXiv:1705.04650 [hep-ex].

[116] CMS Collaboration. «Search for third-generation scalar leptoquarks decaying to a top quark and a $\tau$ lepton at $\sqrt{s} = 13$ TeV.» In: *Eur. Phys. J. C* 78 (2018), p. 707. DOI: 10.1140/epjc/s10052-018-6143-z. arXiv: 1803.02864 [hep-ex].

[117] Albert M Sirunyan et al. «Search for leptoquarks coupled to third-generation quarks in proton-proton collisions at $\sqrt{s} = 13$ TeV.» In: *Phys. Rev. Lett.* 121.24 (2018), p. 241802. DOI: 10.1103/PhysRevLett.121.241802. arXiv: 1809.05558 [hep-ex].

[118] CMS Collaboration. «Search for standard model production of four top quarks with same-sign and multilepton final states in proton–proton collisions at $\sqrt{s} = 13$ TeV.» In: *Eur. Phys. J. C* 78.2 (2018), p. 140. DOI: 10.1140/epjc/s10052-018-5607-5. arXiv: arXiv:1710.10614 [hep-ex].

[119] Ilja Doršner and Admir Greljo. «Leptoquark toolbox for precision collider studies.» In: *JHEP* 05 (2018), p. 126. DOI: 10.1007/JHEP05(2018)126. arXiv: arXiv:1801.07641 [hep-ph].

[120] ATLAS Collaboration. «Search for supersymmetry in events with b-tagged jets and missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector.» In: *JHEP* 11 (2017), p. 195. DOI: 10.1007/JHEP11(2017)195. arXiv: arXiv:1708.09266 [hep-ex].

[121] Richard D. Ball et al. «Parton distributions with LHC data.» In: *Nucl. Phys. B* 867 (2013), p. 244. DOI: 10.1016/j.nuclphysb.2012.10.003. arXiv: arXiv:1207.1303 [hep-ph].

[122] Christoph Borschensky, Michael Krämer, Anna Kulesza, Michelangelo Mangano, Sanjay Padhi, Tilman Plehn, and Xavier Portell. «Squark and gluino production cross sections in pp collisions at $\sqrt{s}$ = 13, 14, 33 and 100 TeV.» In: *Eur. Phys. J. C* 74.12 (2014), p. 3174. DOI: 10.1140/epjc/s10052-014-3174-y. arXiv: arXiv:1407.5066 [hep-ph].

[123] Alexander L. Read. «Presentation of search results: The CL(s) technique.» In: *J. Phys. G* 28 (2002), p. 2693. DOI: 10.1088/0954-3899/28/10/313.

[124] M. Baak, G. J. Besjes, D. Côté, A. Koutsman, J. Lorenz, and D. Short. «HistFitter software framework for statistical data analysis.» In: *Eur. Phys. J. C* 75 (2015), p. 153. DOI: 10.1140/epjc/s10052-015-3327-7. arXiv: arXiv:1410.1280 [hep-ex].

[125] W. Verkerke and D. Kirkb. «RooFit Users Manual v2.07.» In: (). https://roofit.sourceforge.net/docs/RooFit_Users_Manual_2.07-29.pdf.

[126] F. James. «MINUIT Function Minimization and Error Analysis: Reference Manual Version 94.1.» In: (1994).

[127] ATLAS Collaboration. «Prospects for BSM searches at the high-luminosity LHC with the ATLAS detector.» In: *Nucl. Part. Phys. Proc.* 273 (2016), p. 625. DOI: 10.1016/j.nuclphysbps.2015.09.094.

[128] Albert M Sirunyan et al. «Search for singly and pair-produced leptoquarks coupling to third-generation fermions in proton-proton collisions at $\sqrt{s} = 13$ TeV.» In: (Dec. 2020). arXiv: 2012.04178 [hep-ex].

[129] Albert M Sirunyan et al. «Search for pair production of second-generation leptoquarks at $\sqrt{s} = 13$ TeV.» In: *Phys. Rev. D* 99.3 (2019), p. 032014. DOI: 10.1103/PhysRevD.99.032014. arXiv: 1808.05082 [hep-ex].

[130] Ezequiel Alvarez and Manuel Szewc. «Nonresonant leptoquark with multigeneration couplings for $\mu\mu jj$ and $\mu\nu jj$ at the LHC.» In: *Phys. Rev. D* 99.9 (2019), p. 095004. DOI: 10.1103/PhysRevD.99.095004. arXiv: 1811.05944 [hep-ph].

[131] Mark B. Wise and Yue Zhang. «Effective Theory and Simple Completions for Neutrino Interactions.» In: *Phys. Rev. D* 90.5 (2014), p. 053005. DOI: 10.1103/PhysRevD.90.053005. arXiv: 1404.4663 [hep-ph].

[132] Tanumoy Mandal, Subhadip Mitra, and Swapnil Raz. «$R_{D^{(*)}}$ motivated $S_1$ leptoquark scenarios: Impact of interference on the exclusion limits from LHC data.» In: *Phys. Rev. D* 99.5 (2019), p. 055028. DOI: 10.1103/PhysRevD.99.055028. arXiv: 1811.03561 [hep-ph].

[133] Saurabh Bansal, Rodolfo M. Capdevilla, Antonio Delgado, Christopher Kolda, Adam Martin, and Nirmal Raj. «Hunting leptoquarks in monolepton searches.» In: *Phys. Rev. D* 98.1 (2018), p. 015037. DOI: 10.1103/PhysRevD.98.015037. arXiv: 1806.02370 [hep-ph].

[134] Morad Aaboud et al. «Search for squarks and gluinos in final states with jets and missing transverse momentum using 36 $fb^{-1}$ of $\sqrt{s} = 13$ TeV pp collision data with the ATLAS detector.» In: *Phys. Rev. D* 97.11 (2018), p. 112001. DOI: 10.1103/PhysRevD.97.112001. arXiv: 1712.02332 [hep-ex].

[135] Morad Aaboud et al. «Search for scalar leptoquarks in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS experiment.» In: *New J. Phys.* 18.9 (2016), p. 093016. DOI: 10.1088/1367-2630/18/9/093016. arXiv: 1605.06035 [hep-ex].

[136] Damir Beˇ cirević, Nejc Koˇ snik, Olcyr Sumensari, and Renata Zukanovich Funchal. «Palatable Leptoquark Scenarios for Lepton Flavor Violation in Exclusive $b \to s\ell_1\ell_2$ modes.» In: *JHEP* 11 (2016), p. 035. DOI: 10.1007/JHEP11(2016)035. arXiv: 1608.07583 [hep-ph].

[137] A.D. Martin, W.J. Stirling, R.S. Thorne, and G. Watt. «Parton distributions for the LHC.» In: *Eur. Phys. J. C* 63 (2009), pp. 189–285. DOI: 10.1140/epjc/s10052-009-1072-5. arXiv: 0901.0002 [hep-ph].

[138] M. Kramer, T. Plehn, M. Spira, and P.M. Zerwas. «Pair production of scalar leptoquarks at the CERN LHC.» In: *Phys. Rev. D* 71 (2005), p. 057503. DOI: 10.1103/PhysRevD.71.057503. arXiv: hep-ph/0411038.

[139] Tanumoy Mandal, Subhadip Mitra, and Satyajit Seth. «Pair Production of Scalar Leptoquarks at the LHC to NLO Parton Shower Accuracy.» In: *Phys. Rev. D* 93.3 (2016), p. 035018. DOI: 10.1103/PhysRevD.93.035018. arXiv: 1506.07369 [hep-ph].

[140] Ferruccio Feruglio, Paride Paradisi, and Andrea Pattori. «On the Importance of Electroweak Corrections for B Anomalies.» In: *JHEP* 09 (2017), p. 061. DOI: 10.1007/JHEP09(2017)061. arXiv: 1705.00929 [hep-ph].

[141] Martín González-Alonso, Jorge Martin Camalich, and Kin Mimouni. «Renormalization-group evolution of new physics contributions to (semi)leptonic meson decays.» In: *Phys. Lett. B* 772 (2017), pp. 777–785. DOI: 10.1016/j.physletb.2017.07.003. arXiv: 1706.00410 [hep-ph].

[142] Ferruccio Feruglio, Paride Paradisi, and Olcyr Sumensari. «Implications of scalar and tensor explanations of $R_{D^{(*)}}$.» In: *JHEP* 11 (2018), p. 191. DOI: 10.1007/JHEP11(2018)191. arXiv: 1806.10155 [hep-ph].

[143] J. Grygier et al. «Search for $B \to h\nu\bar{\nu}$ decays with semileptonic tagging at Belle.» In: *Phys. Rev. D* 96.9 (2017). [Addendum: Phys.Rev.D 97, 099902 (2018)], p. 091101. DOI: 10.1103/PhysRevD.96.091101. arXiv: 1702.03224 [hep-ex].

[144] M. Tanabashi et al. «Review of Particle Physics.» In: *Phys. Rev. D* 98.3 (2018), p. 030001. DOI: 10.1103/PhysRevD.98.030001.

[145] Ezequiel Alvarez, Leandro Da Rold, Mariel Estevez, and Jernej F. Kamenik. «Measuring $|V_{td}|$ at the LHC.» In: *Phys. Rev. D* 97.3 (2018), p. 033002. DOI: 10.1103/PhysRevD.97.033002. arXiv: 1709.07887 [hep-ph].

[146] Vardan Khachatryan et al. «Measurement of the ratio $\mathcal{B}(t \to Wb)/\mathcal{B}(t \to Wq)$ in pp collisions at $\sqrt{s} = 8$ TeV.» In: *Phys. Lett. B* 736 (2014), pp. 33–57. DOI: 10.1016/j.physletb.2014.06.076. arXiv: 1404.2292 [hep-ex].

[147] Ezequiel Alvarez, Aurelio Juste, Manuel Szewc, and Tamara Vazquez Schroeder. «Topping-up multilepton plus b-jets anomalies at the LHC with a $Z'$ boson.» In: (Nov. 2020). arXiv: 2011.06514 [hep-ph].

[148] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. «Asymptotic formulae for likelihood-based tests of new physics.» In: *Eur. Phys. J. C* 71 (2011). [Erratum: Eur. Phys. J. C 73 (2013) 2501], p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: arXiv:1007.1727 [physics.data-an].

[149] ATLAS Collaboration. «Measurement of the $t\bar{t}Z$ and $t\bar{t}W$ cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector.» In: *Phys. Rev. D* 99.7 (2019), p. 072009. DOI: 10.1103/PhysRevD.99.072009. arXiv: arXiv:1901.03584 [hep-ex].

[150] CMS Collaboration. «Measurement of the cross section for top quark pair production in association with a W or Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV.» In: *JHEP* 08 (2018), p. 011. DOI: 10.1007/JHEP08(2018)011. arXiv: arXiv:1711.02547 [hep-ex].

[151] ATLAS Collaboration. «Analysis of $t\bar{t}H$ and $t\bar{t}W$ production in multilepton final states with the ATLAS detector.» In: *ATLAS-CONF-2019-045* (2019). https://cds.cern.ch/record/2693930.

[152] CMS Collaboration. «Measurement of the Higgs boson production rate in association with top quarks in final states with electrons, muons, and hadronically decaying tau leptons at $\sqrt{s} = 13$ TeV.» In: (Nov. 2020). arXiv: arXiv:2011.03652 [hep-ex].

[153] D. de Florian et al. «Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector.» In: 2/2017 (Oct. 2016). DOI: 10.23731/CYRM-2017-002. arXiv: arXiv:1610.07922 [hep-ph].

[154] Albert M Sirunyan et al. «Search for new physics in top quark production with additional leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV using effective field theory.» In: (Dec. 2020). arXiv: arXiv:2012.04120 [hep-ex].

[155] Giovanni Banelli, Ennio Salvioni, Javi Serra, Tobias Theil, and Andreas Weiler. «The Present and Future of Four Tops.» In: (Oct. 2020). arXiv: arXiv:2010.05915 [hep-ph].

[156] L.D. Landau. «On the angular momentum of a system of two photons.» In: *Dokl. Akad. Nauk SSSR* 60.2 (1948), p. 207. DOI: 10.1016/B978-0-08-010586-4.50070-5.

[157] Chen-Ning Yang. «Selection Rules for the Dematerialization of a Particle Into Two Photons.» In: *Phys. Rev.* 77 (1950), p. 242. DOI: 10.1103/PhysRev.77.242.

[158] V. Pleitez. «The angular momentum of two massless fields revisited.» In: (Aug. 2015). arXiv: arXiv:1508.01394 [hep-ph].

[159] ATLAS Collaboration. «Search for Scalar Diphoton Resonances in the Mass Range $65 - 600$ GeV with the ATLAS Detector in pp Collision Data at $\sqrt{s} = 8$ TeV.» In: *Phys. Rev. Lett.* 113.17 (2014), p. 171801. DOI: 10.1103/PhysRevLett.113.171801. arXiv: arXiv:1407.6583 [hep-ex].

[160] ATLAS Collaboration. «Search for new phenomena in high-mass diphoton final states using 37 $\text{fb}^{-1}$ of proton–proton collisions collected at $\sqrt{s} = 13$ TeV with the ATLAS detector.» In: *Phys. Lett. B* 775 (2017), p. 105. DOI: 10.1016/j.physletb.2017.10.039. arXiv: arXiv:1707.04147 [hep-ex].

[161] ATLAS Collaboration. «Search for resonances in the 65 to 110 GeV diphoton invariant mass range using 80 $\text{fb}^{-1}$ of pp collisions collected at $\sqrt{s} = 13$ TeV with the ATLAS detector.» In: *ATLAS-CONF-2018-025* (2018). https://cds.cern.ch/record/2628760.

[162] CMS Collaboration. «Search for diphoton resonances in the mass range from 150 to 850 GeV in pp collisions at $\sqrt{s} = 8$ TeV.» In: *Phys. Lett. B* 750 (2015), p. 494. DOI: 10.1016/j.physletb.2015.09.062. arXiv: arXiv:1506.02301 [hep-ex].

[163] CMS Collaboration. «Search for a standard model-like Higgs boson in the mass range between 70 and 110 GeV in the diphoton final state in proton-proton collisions at $\sqrt{s} = 8$ and 13 TeV.» In: *Phys. Lett. B* 793 (2019), p. 320. DOI: 10.1016/j.physletb.2019.03.064. arXiv: arXiv:1811.08459 [hep-ex].

[164] Stefan Buddenbrock, Alan S. Cornell, Yaquan Fang, Abdualazem Fadol Mohammed, Mukesh Kumar, Bruce Mellado, and Kehinde G. Tomiwa. «The emergence of multi-lepton anomalies at the LHC and their compatibility with new physics at the EW scale.» In: *JHEP* 10 (2019), p. 157. DOI: 10.1007/JHEP10(2019)157. arXiv: arXiv:1901.05300 [hep-ph].

[165] Wei-Shu Hou, Ting-Hsiang Hsu, and Tanmoy Modak. «Constraining the $t \to u$ flavor changing neutral Higgs coupling at the LHC.» In: *Phys. Rev. D* 102.5 (2020), p. 055006. DOI: 10.1103/PhysRevD.102.055006. arXiv: arXiv:2008.02573 [hep-ph].

[166] Patrick J. Fox, Ian Low, and Yue Zhang. «Top-philic Z′ forces at the LHC.» In: *JHEP* 03 (2018), p. 074. DOI: 10.1007/JHEP03(2018)074. arXiv: arXiv:1801.03505 [hep-ph].

[167] Ezequiel Alvarez, Aurelio Juste, and Rosa María Sandá Seoane. «Four-top as probe of light top-philic New Physics.» In: *JHEP* 12 (2019), p. 080. DOI: 10.1007/JHEP12(2019)080. arXiv: arXiv:1910.09581 [hep-ph].

[168] P. Ko, Yuji Omura, and Chaehyun Yu. «Top $A_{FB}$ at the Tevatron vs. Charge Asymmetry at the LHC in Chiral U(1) Flavor Models with Flavored Higgs Doublets.» In: *Eur. Phys. J. C* 73.1 (2013), p. 2269. DOI: 10.1140/epjc/s10052-012-2269-6. arXiv: arXiv:1205.0407 [hep-ph].

[169] Sungwoong Cho, P. Ko, Jungil Lee, Yuji Omura, and Chaehyun Yu. «Top FCNC induced by a Z′ boson.» In: *Phys. Rev. D* 101.5 (2020), p. 055015. DOI: 10.1103/PhysRevD.101.055015. arXiv: arXiv:1910.05925 [hep-ph].

[170] Ezequiel Alvarez and Estefania Coluccio Leskow. «A charged Z′ to conciliate the apparent disagreement between top-antitop Tevatron forward-backward asymmetry and LHC charge asymmetry.» In: *Phys. Rev. D* 86 (2012), p. 114034. DOI: 10.1103/PhysRevD.86.114034. arXiv: arXiv:1209.4354 [hep-ph].

[171] Nathaniel Craig, Jan Hajer, Ying-Ying Li, Tao Liu, and Hao Zhang. «Heavy Higgs bosons at low $\tan\beta$: from the LHC to 100 TeV.» In: *JHEP* 01 (2017), p. 018. DOI: 10.1007/JHEP01(2017)018. arXiv: arXiv:1605.08744 [hep-ph].

[172] Wei-Shu Hou, Masaya Kohda, and Tanmoy Modak. «Search for $tZ'$ associated production induced by $tcZ'$ couplings at the LHC.» In: *Phys. Rev. D* 96.1 (2017), p. 015037. DOI: 10.1103/PhysRevD.96.015037. arXiv: arXiv:1702.07275 [hep-ph].

[173] S. Frixione, V. Hirschi, D. Pagani, H. S. Shao, and M. Zaro. «Electroweak and QCD corrections to top-pair hadroproduction in association with heavy bosons.» In: *JHEP* 06 (2015), p. 184. DOI: 10.1007/JHEP06(2015)184. arXiv: arXiv:1504.03446 [hep-ph].

[174] Alessandro Broggio, Andrea Ferroglia, Rikkert Frederix, Davide Pagani, Benjamin D. Pecjak, and Ioannis Tsinikos. «Top-quark pair hadroproduction in association with a heavy boson at NLO+NNLL including EW corrections.» In: *JHEP* 08 (2019), p. 039. DOI: 10.1007/JHEP08(2019)039. arXiv: arXiv:1907.04343 [hep-ph].

[175] Anna Kulesza, Leszek Motyka, Daniel Schwartländer, Tomasz Stebel, and Vincent Theeuwes. «Associated top quark pair production with a heavy boson: differential cross sections at NLO+NNLL accuracy.» In: *Eur. Phys. J. C* 80.5 (2020), p. 428. DOI: 10.1140/epjc/s10052-020-7987-6. arXiv: arXiv:2001.03031 [hep-ph].

[176] Rikkert Frederix and Ioannis Tsinikos. «Subleading EW corrections and spin-correlation effects in $t\bar{t}W$ multi-lepton signatures.» In: *Eur. Phys. J. C* 80.9 (2020), p. 803. DOI: 10.1140/epjc/s10052-020-8388-6. arXiv: arXiv:2004.09552 [hep-ph].

[177] Giuseppe Bevilacqua, Huan-Yu Bi, Heribertus Bayu Hartanto, Manfred Kraus, and Malgorzata Worek. «The simplest of them all: $t\bar{t}W^{\pm}$ at NLO accuracy in QCD.» In: *JHEP* 08 (2020), p. 043. DOI: 10.1007/JHEP08(2020)043. arXiv: arXiv:2005.09427 [hep-ph].

[178] Ansgar Denner and Giovanni Pelliccioli. «NLO QCD corrections to off-shell $t\bar{t}W^{+}$ production at the LHC.» In: (July 2020). arXiv: arXiv:2007.12089 [hep-ph].

[179] Stefan von Buddenbrock, Richard Ruiz, and Bruce Mellado. «Anatomy of inclusive $t\bar{t}W$ production at hadron colliders.» In: (Aug. 2020). arXiv: arXiv:2009.00032 [hep-ph].

[180] Peter Skands, Stefano Carrazza, and Juan Rojo. «Tuning PYTHIA 8.1: the Monash 2013 Tune.» In: *Eur. Phys. J. C* 74.8 (2014), p. 3024. DOI: 10.1140/epjc/s10052-014-3024-y. arXiv: arXiv:1404.5630 [hep-ph].

[181] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. «The anti-$k_t$ jet clustering algorithm.» In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[182] Marco Guzzi and Nikolaos Kidonakis. «$tZ'$ production at hadron colliders.» In: *Eur. Phys. J. C* 80.5 (2020), p. 467. DOI: 10.1140/epjc/s10052-020-8028-1. arXiv: arXiv:1904.10071 [hep-ph].

[183] R. D. Cousins. «Generalization of Chisquare Goodness-of-Fit Test for Binned Data Using Saturated Models, with Application to Histograms.» In: (). http://www.physics.ucla.edu/~cousins/stats/cousins_saturated.pdf.

[184] Eugene Golowich, JoAnne Hewett, Sandip Pakvasa, and Alexey A. Petrov. «Implications of $D^0$ - $\bar{D}^0$ Mixing for New Physics.» In: *Phys. Rev. D* 76 (2007), p. 095009. DOI: 10.1103/PhysRevD.76.095009. arXiv: arXiv:0705.3650 [hep-ph].

[185] J. I. Aranda, F. Ramirez-Zavaleta, J. J. Toscano, and E. S. Tututi. «Bounding the $Z'tc$ coupling from $D^0 - \overline{D^0}$ mixing and single top production at the ILC.» In: *J. Phys. G* 38 (2011), p. 045006. DOI: 10.1088/0954-3899/38/4/045006. arXiv: arXiv:1007.3326 [hep-ph].

[186] M. Artuso et al. «Improved measurement of B (D+ —> mu+ nu) and the pseudoscalar decay constant f(D+).» In: *Phys. Rev. Lett.* 95 (2005), p. 251801. DOI: 10.1103/PhysRevLett.95.251801. arXiv: hep-ex/0508057.

[187] Rajan Gupta, Tanmoy Bhattacharya, and Stephen R. Sharpe. «Matrix elements of four fermion operators with quenched Wilson fermions.» In: *Phys. Rev. D* 55 (1997), pp. 4036–4054. DOI: 10.1103/PhysRevD.55.4036. arXiv: hep-lat/9611023.

[188] Enrico Franco, Satoshi Mishima, and Luca Silvestrini. «The Standard Model confronts CP violation in $D^0 \to \pi^+\pi^-$ and $D^0 \to K^+K^-$.» In: *JHEP* 05 (2012), p. 140. DOI: 10.1007/JHEP05(2012)140. arXiv: arXiv:1203.3131 [hep-ph].

[189] Gino Isidori, Jernej F. Kamenik, Zoltan Ligeti, and Gilad Perez. «Implications of the LHCb Evidence for Charm CP Violation.» In: *Phys. Lett. B* 711 (2012), p. 46. DOI: 10.1016/j.physletb.2012.03.046. arXiv: arXiv:1111.4987 [hep-ph].

[190] ATLAS Collaboration. «Measurement of the $t\bar{t}$ production cross-section in the lepton+jets channel at $\sqrt{s} = 13$ TeV with the ATLAS experiment.» In: (June 2020). arXiv: arXiv:2006.13076 [hep-ex].

[191] CMS Collaboration. «Measurement of the top quark pair production cross section in proton-proton collisions at $\sqrt{s} = 13$ TeV.» In: *Phys. Rev. Lett.* 116.5 (2016), p. 052002. DOI: 10.1103/PhysRevLett.116.052002. arXiv: arXiv:1510.05302 [hep-ex].

[192] CMS Collaboration. «Search for physics beyond the standard model in events with two leptons of same sign, missing transverse momentum, and jets in proton-proton collisions at $\sqrt{s} = 13$ TeV.» In: *Eur. Phys. J. C* 77.9 (2017), p. 578. DOI: 10.1140/epjc/s10052-017-5079-z. arXiv: arXiv:1704.07323 [hep-ex].

[193] ATLAS Collaboration. «Search for new phenomena in events with same-charge leptons and b-jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector.» In: *JHEP* 12 (2018), p. 039. DOI: 10.1007/JHEP12(2018)039. arXiv: arXiv:1807.11883 [hep-ex].

[194] ATLAS Collaboration. «Search for resonant top plus jet production in $t\bar{t}$ + jets events with the ATLAS detector in pp collisions at $\sqrt{s} = 7$ TeV.» In: *Phys. Rev. D* 86 (2012), p. 091103. DOI: 10.1103/PhysRevD.86.091103. arXiv: arXiv:1209.6593 [hep-ex].

[195] CDF Collaboration. «Search for a heavy particle decaying to a top quark and a light quark in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV.» In: *Phys. Rev. Lett.* 108 (2012), p. 211805. DOI: 10.1103/PhysRevLett.108.211805. arXiv: arXiv:1203.3894 [hep-ex].

[196] Nathaniel Craig, Jan Hajer, Ying-Ying Li, Tao Liu, and Hao Zhang. «Heavy Higgs bosons at low $\tan\beta$: from the LHC to 100 TeV.» In: *JHEP* 01 (2017), p. 018. DOI: 10.1007/JHEP01(2017)018. arXiv: arXiv:1605.08744 [hep-ph].

[197] Andreas Hocker et al. «TMVA - Toolkit for Multivariate Data Analysis.» In: (Mar. 2007). arXiv: arXiv:physics/0703039.

[198] Ezequiel Alvarez, Federico Lamagna, and Manuel Szewc. «Topic Model for four-top at the LHC.» In: *JHEP* 01 (2020), p. 049. DOI: 10.1007/JHEP01(2020)049. arXiv: 1911.09699 [hep-ph].

[199] Ezequiel Alvarez et al. *Keras*. https://github.com/ManuelSzewc/Topic-Model-for-four-top-at-the-LHC. 2019.

[200] Albert M Sirunyan et al. «Search for the Production of Four Top Quarks in the Single-Lepton and Opposite-Sign Dilepton Final States in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV.» In: (2019). arXiv: 1906.02805 [hep-ex].

[201] Rikkert Frederix, Davide Pagani, and Marco Zaro. «Large NLO corrections in $t\bar{t}W^{\pm}$ and $t\bar{t}t\bar{t}$ hadroproduction from supposedly subleading EW contributions.» In: (2017). arXiv: 1711.02116 [hep-ph].

[202] Patrick T. Komiske, Eric M. Metodiev, and Jesse Thaler. «An operational definition of quark and gluon jets.» In: *Journal of High Energy Physics* 2018.11 (Nov. 2018), p. 59. ISSN: 1029-8479. DOI: 10.1007/JHEP11(2018)059. URL: https://doi.org/10.1007/JHEP11(2018)059.

[203] Julian Katz-Samuels, Gilles Blanchard, and Clayton Scott. «Decontamination of Mutual Contamination Models.» In: (2017). arXiv: 1710.01167 [stat.ML].

[204] Michel Marie Deza and Elena Deza. *Encyclopedia of Distances*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 1–583. ISBN: 978-3-642-00234-2. DOI: 10.1007/978-3-642-00234-2_1. URL: https://doi.org/10.1007/978-3-642-00234-2_1.

[205] François Chollet et al. *Keras*. https://keras.io. 2015.

[206] ATLAS Collaboration. *Optimisation of the ATLAS* b*-tagging performance for the 2016 LHC Run*. Tech. rep. ATL-PHYS-PUB-2016-012. Geneva: CERN, June 2016. URL: http://cds.cern.ch/record/2160731.

[207] David M. Blei. «Probabilistic Topic Models.» In: *Commun. ACM* 55.4 (Apr. 2012), 77–84. ISSN: 0001-0782. DOI: 10.1145/2133806.2133826. URL: https://doi.org/10.1145/2133806.2133826.

[208] David M. Blei, Andrew Y. Ng, Michael I. Jordan, and John Lafferty. «Latent dirichlet allocation.» In: *Journal of Machine Learning Research* 3 (2003), p. 2003.

[209] Thomas L. Griffiths and Mark Steyvers. «Finding scientific topics.» In: *Proceedings of the National Academy of Sciences* 101.suppl 1 (2004), pp. 5228–5235. ISSN: 0027-8424. DOI: 10.1073/pnas.0307752101.

[210] Jonathan K. Pritchard, Matthew Stephens, and Peter Donnelly. «Inference of Population Structure Using Multilocus Genotype Data.» In: *Genetics* 155.2 (2000), pp. 945–959. ISSN: 0016-6731. eprint: http://www.genetics.org/content/155/2/945.full.pdf. URL: http://www.genetics.org/content/155/2/945.

[211] Thomas Hofmann. «Probabilistic Latent Semantic Analysis.» In: *Proc. of Uncertainty in Artificial Intelligence, UAI-99* (1999), pp. 289–296.

[212] Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. «Indexing by latent semantic analysis.» In: *JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE* 41.6 (1990), pp. 391–407.

[213] Kamal Nigam, Andrew Kachites Mccallum, Sebastian Thrun, and Tom Mitchell. «Text Classification from Labeled and Unlabeled Documents using EM.» In: *MACHINE LEARNING* (1999), pp. 103–134.

[214] David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. «Variational Inference: A Review for Statisticians.» In: *Journal of the American Statistical Association* 112.518 (Feb. 2017), 859–877. ISSN: 1537-274X. DOI: 10.1080/01621459.2017.1285773. URL: http://dx.doi.org/10.1080/01621459.2017.1285773.

[215] J.A. Aguilar-Saavedra. «Stealth multiboson signals.» In: *Eur. Phys. J. C* 77.10 (2017), p. 703. DOI: 10.1140/epjc/s10052-017-5289-4. arXiv: 1705.07885 [hep-ph].

[216] J.A. Aguilar-Saavedra and F.R. Joaquim. «The minimal stealth boson: models and benchmarks.» In: *JHEP* 10 (2019), p. 237. DOI: 10.1007/JHEP10(2019)237. arXiv: 1905.12651 [hep-ph].

[217] B.M. Dillon, D.A. Faroughy, J.F. Kamenik, and M. Szewc. «Learning the latent structure of collider events.» In: *JHEP* 10 (2020), p. 206. DOI: 10.1007/JHEP10(2020)206. arXiv: 2005.12319 [hep-ph].

[218] Barry Dillon et al. *Keras*. https://github.com/barrydillon89/LDA-jet-substructure. 2019.

[219] E. Hewitt and L. J. Savage. «Symmetric measures on Cartesian products.» In: *Transactions of the American Mathematical Society* 80 (1955), 470–501.

[220] Arjun Gupta and Saralees Nadarajah. «Handbook of Beta Distribution and Its Applications.» In: *CRC Press* (Jan. 2004).

[221] Matthew D. Hoffman, David M. Blei, and Francis Bach. «Online Learning for Latent Dirichlet Allocation.» In: *Proceedings of the 23rd International Conference on Neural Information Processing Systems - Volume 1*. NIPS'10 (2010), pp. 856–864. URL: http://dl.acm.org/citation.cfm?id=2997189.2997285.

[222] S. Kullback and R.A. Leibler. «On information and sufficiency.» In: *Annals of Mathematical Statistics* 22.1 (1951), 79–86.

[223] Radim Řehůřek and Petr Sojka. «Software Framework for Topic Modelling with Large Corpora.» English. In: *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks* (May 2010). http://is.muni.cz/publication/884893/en, pp. 45–50.

[224] T. P. Minka. «Estimating a Dirichlet distribution.» In: (2012). https://tminka.github.io/papers/dirichlet/minka-dirichlet.pdf.

[225] S. Catani, Yuri L. Dokshitzer, M. H. Seymour, and B. R. Webber. «Longitudinally invariant $K_t$ clustering algorithms for hadron hadron collisions.» In: *Nucl. Phys.* B406 (1993), pp. 187–224. DOI: 10.1016/0550-3213(93)90166-M.

[226] Stephen D. Ellis and Davison E. Soper. «Successive combination jet algorithm for hadron collisions.» In: *Phys. Rev.* D48 (1993), pp. 3160–3166. DOI: 10.1103/PhysRevD.48.3160. arXiv: hep-ph/9305266 [hep-ph].

[227] Yuri L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber. «Better jet clustering algorithms.» In: *JHEP* 08 (1997), p. 001. DOI: 10.1088/1126-6708/1997/08/001. arXiv: hep-ph/9707323 [hep-ph].

[228] M. Wobisch and T. Wengler. «Hadronization corrections to jet cross-sections in deep inelastic scattering.» In: *Monte Carlo generators for HERA physics. Proceedings, Workshop, Hamburg, Germany, 1998-1999*. 1998, pp. 270–279. arXiv: hep-ph/9907280 [hep-ph].

[229] Tilman Plehn, Michael Spannowsky, Michihisa Takeuchi, and Dirk Zerwas. «Stop Reconstruction with Tagged Tops.» In: *JHEP* 10 (2010), p. 078. DOI: 10.1007/JHEP10(2010)078. arXiv: 1006.2833 [hep-ph].

[230] Frédéric A. Dreyer, Gavin P. Salam, and Grégory Soyez. «The Lund Jet Plane.» In: *JHEP* 12 (2018), p. 064. DOI: 10.1007/JHEP12(2018)064. arXiv: 1807.04758 [hep-ph].

[231] Tilman Plehn, Gavin P. Salam, and Michael Spannowsky. «Fat Jets for a Light Higgs.» In: *Phys. Rev. Lett.* 104 (2010), p. 111801. DOI: 10.1103/PhysRevLett.104.111801. arXiv: 0910.5472 [hep-ph].

[232] Thorsten Chwalek. «Measurement of the $W^-$ boson helicity fractions in top-quark decays at CDF.» In: *42nd Rencontres de Moriond on Electroweak Interactions and Unified Theories*. May 2007, pp. 495–489. arXiv: 0705.2966 [hep-ex].

[233] David E. Kaplan, Keith Rehermann, Matthew D. Schwartz, and Brock Tweedie. «Top Tagging: A Method for Identifying Boosted Hadronically Decaying Top Quarks.» In: *Phys. Rev. Lett.* 101 (2008), p. 142001. DOI: 10.1103/PhysRevLett.101.142001. arXiv: 0806.0848 [hep-ph].

[234] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. «FastJet User Manual.» In: *Eur. Phys. J.* C72 (2012), p. 1896. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph].

[235] Anja Butter et al. «The Machine Learning Landscape of Top Taggers.» In: *SciPost Phys.* 7 (2019). Ed. by Gregor Kasieczka and Tilman Plehn, p. 014. DOI: 10.21468/SciPostPhys.7.1.014. arXiv: 1902.09914 [hep-ph].

[236] Jerzy Neyman and Egon Sharpe Pearson. «On the Problem of the Most Efficient Tests of Statistical Hypotheses.» In: *Phil. Trans. Roy. Soc. Lond. A* 231.694-706 (1933), pp. 289–337. DOI: 10.1098/rsta.1933.0009.

[237] Chong Wang, John Paisley, and David Blei. «Online Variational Inference for the Hierarchical Dirichlet Process.» In: Proceedings of Machine Learning Research 15 (2011). Ed. by Geoffrey Gordon, David Dunson, and Miroslav Dudík, pp. 752–760. URL: http://proceedings.mlr.press/v15/wang11a.html.

# LIST OF PUBLICATIONS


[1] Ezequiel Alvarez, Leandro Da Rold, Aurelio Juste, Manuel Szewc, and Tamara Vazquez Schroeder. «A composite pNGB leptoquark at the LHC.» In: *JHEP* 12 (2018), p. 027. DOI: 10.1007/JHEP12(2018)027. arXiv: 1808.02063 [hep-ph].

[2] Ezequiel Alvarez and Manuel Szewc. «Nonresonant leptoquark with multigeneration couplings for μμjj and μνjj at the LHC.» In: *Phys. Rev. D* 99.9 (2019), p. 095004. DOI: 10.1103/PhysRevD.99.095004. arXiv: 1811.05944 [hep-ph].

[3] Ezequiel Alvarez, Federico Lamagna, and Manuel Szewc. «Topic Model for four-top at the LHC.» In: *JHEP* 01 (2020), p. 049. DOI: 10.1007/JHEP01(2020)049. arXiv: 1911.09699 [hep-ph].

[4] B.M. Dillon, D.A. Faroughy, J.F. Kamenik, and M. Szewc. «Learning the latent structure of collider events.» In: *JHEP* 10 (2020), p. 206. DOI: 10.1007/JHEP10(2020)206. arXiv: 2005.12319 [hep-ph].

[5] Ezequiel Alvarez, Aurelio Juste, Manuel Szewc, and Tamara Vazquez Schroeder. «Topping-up multilepton plus b-jets anomalies at the LHC with a Z′ boson.» In: (Nov. 2020). arXiv: 2011.06514 [hep-ph].

[6] Gregor Kasieczka et al. «The LHC Olympics 2020: A Community Challenge for Anomaly Detection in High Energy Physics.» In: (Jan. 2021). arXiv: 2101.08320 [hep-ph].

[1] Ezequiel Alvarez, Federico Lamagna, Cesar Miquel, and Manuel Szewc. «Intelligent Arxiv: Sort daily papers by learning users topics preference.» In: (Feb. 2020). arXiv: 2002.02460 [cs.LG].

[2] Ezequiel Alvarez, Federico Lamagna, and Manuel Szewc. «A Machine Learning alternative to placebo-controlled clinical trials upon new diseases: A primer.» In: (Mar. 2020). arXiv: 2003.12454 [q-bio.QM].

[3] Ezequiel Alvarez, Leandro Da Rold, Federico Lamagna, and Manuel Szewc. «Containing COVID-19 outbreaks using a Firewall.» In: (Aug. 2020). arXiv: 2008.12636 [physics.soc-ph].