ANDREAS SØGAARD

BOOSTED BOSONS AND WAVELETS

BOOSTED BOSONS AND WAVELETS

ANDREAS SØGAARD

NIELS BOHR INSTITUTE,
UNIVERSITY OF COPENHAGEN

SUPERVISOR	Dr. T. Petersen
REFEREE	Dr. C. Doglioni (LUND UNIVERSITY)

UNIVERSITY OF
COPENHAGEN

Questions of science,

science and progress,

do not speak as loud

as my heart.

— Coldplay, *The Scientist*

Contents

Abstract

For the LHC Run 2 and beyond, experiments are pushing both the energy and the intensity frontier so the need for robust and efficient pile-up mitigation tools becomes ever more pressing. Several methods exist, relying on uniformity of pile-up, local correlations of charged to neutral particles, and parton shower shapes, all in $y - \phi$ space. Wavelets are presented as tools for pile-up removal, utilising their ability to encode position and frequency information simultaneously. This allows for the separation of individual hadron collision events by angular scale and thus for subtracting of soft, diffuse/wide-angle contributions while retaining the hard, small-angle components from the hard event. Wavelet methods may utilise the same assumptions as existing methods, the difference being the underlying, novel representation. Several wavelet methods are proposed and their effect studied in simple toy simulation under conditions relevant for the LHC Run 2. One full pile-up mitigation tool ('wavelet analysis') is optimised and its impact on a few jet kinematic variables assessed in both toy and official 13 TeV ATLAS MC. Finally, a mock search for new resonances in the semi-leptonic *WW* channel is presented, focusing on the sensitivity improvements achievable using the wavelet analysis. It is found that jet energy bias may be removed and resolution improved by $\mathcal{O}(50\%)$ for $p_\perp = 300$ GeV jets at $\langle\mu\rangle = 40$. Similarly, jet mass sensitivity for boosted boson jets may be improved by $\mathcal{O}(100\%)$ under similar conditions. The latter has the effect of increasing the semi-leptonic diboson search sensitivity at $\langle\mu\rangle \approx 25$ by upwards of 10% for resonance masses relevant for Run 2. Therefore, analyses at ATLAS—as well as e.g. CMS and ALICE—may immediately benefit from employing wavelet-based methods, both for searches and for other specialised tasks. The impact of using wavelet analyses only increases with $\langle\mu\rangle$, underlining their promise at current and future hadron collider experiments.

Disclaimer

Some of the results presented in this thesis, be they plots or figures, are based on official ATLAS simulated data from the mc14 simulation project. These are the product of the author's own work and have *not* been approved by the ATLAS Collaboration, and therefore they are labeled as 'Work in progress'. Any replication of these results should clearly reflect this fact.

Acknowledgement

During the past year, I have had the privilege to work on a topic which has been both interesting, demanding, and extremely fulfilling. I have had the freedom to work independently and to pursue questions that I found interesting; all thanks to my supervisor, Troels Petersen. I am deeply grateful to Troels for supporting me in joining the CERN Summer Student programme; for introducing me to this fascinating topic and for generously sharing his expertise; for having faith in me and for relentlessly encouraging me along the way; and, above all, for simply being a great guy. Thank you.

I would like to extend my sincere thanks to Prof. Peter Hansen and the rest of the High Energy Physics group at the Niels Bohr Institute, for being truly welcoming, for offering me a great opportunity, and equally great facilities, for carrying out the present thesis work. Specifically, I would like to thank Alejandro 'Alex' Alonso, Craig Wiglesworth, and Geert Jan Besjes for sharing their office with me for the past year, and for always being ready to help out with any technical problems I might have had. Similarly, I want to thank James Monk for producing the MC samples used in Chap. 3 as well as for valuable discussions. Also, I am truly thankful for the group's supporting me in participating in both the ATLAS Grid Tutorial in Geilo, Norway, as well as the Nordic Winter School on Cosmology and Particle Physics in Skeikampen, Norway.

Finally, and most importantly, I could not have done any of this, had it not been for the unwavering support of my girlfriend, Louise. *Du er min klippe i et stormfuldt hav.*

Introduction

> '*But why are such terrific efforts made just to find new particles?*' *asked Mr Tompkins.*
>
> '*Well, this is science,*' *replied the professor,* '*the attempt of the human mind to understand everything around us, be it giant stellar galaxies, microscopic bacteria, or these elementary particles. It is interesting and exciting, and that is why we are doing it.*'
>
> — From *Mr Tompkins Tastes a Japanese Meal*, by George Gamow

■ Overview

For the last half century, physicists have incrementally been putting together the pieces of a large puzzle: the Standard Model of particle physics. Piece by piece, with the inclusion of new particles and forces, through theoretical predictions and experimental measurements, this model has grown and with it our understanding of the world around us. The last piece was added in 2012 by the discovery of what appears to be the Higgs boson, the manifestation of the the long sought-for mechanism responsible for giving masses to other particles. Even though the Standard Model is a marvelous achievement, with countless of experimental successes on record, we know that it is incomplete.

To mend it, numerous theories have been proposed, all of which bring with them new signatures, not predicted by the Standard Model, usually in the form of new particles. Experiments are build to search for these new particles, which show themselves as excesses—in some observable—over what we would expect, had the Standard Model been complete. When we find excesses which cannot be explained by the Standard Model alone we claim a discovery, as was the case for the Higgs boson. If no such significant excesses are found, we can set ever tighter exclusion limits, drawing incremental lines in the sand where we know that nothing new is to be found. As these limits are continuously pushed upwards, physicists must build larger experiments and develop better methods to look further into the unknown.

■ **Outline of thesis**

Thesis thesis attempts to give an accessible and self-contained presentation of the development of a set of techniques which may help increase the sensitivity in such searches for new physics. Specifically, these techniques are based on the mathematical tool of wavelets, and are intended to remove pile-up and improve the reconstruction of hadronically decaying vector bosons, and therefore aid in searches for new particles coupling to these. The thesis is divided into four chapters:

The first chapter covers the necessary background information needed for the remainder of the thesis. It introduces the basics of theoretical particle physics, how this physics manifests itself at hadron colliders, the experimental apparatus designed to measure these collisions, and the ways in which meaningful physics objects are reconstructed from the bare measurements. This chapter serves as an introduction to readers who are not familiar with particle physics, and thus may be omitted by reader who are, to whom the contents should hold no surprises.

The second chapter deals with the concept of wavelets. It briefly presents wavelets in and of themselves and then proceeds to discuss how they might be of use in the context of hadron collions, specifically as a pile-up mitigation tool. Finally, as set of specific wavelet-based methods are presented and motivated.

The third chapter is devoted to the study of the proposed wavelet methods in primitive, simulated data. Specifically, it describes how wavelets behave in simulation and studies the impact of an optimised wavelet analysis on a few jet observables.

The fourth chapter uses realistic simulation to investigate in detail the improvements to jet observables made possible by the optimised wavelet analysis. The thesis concludes with a simple example of a search for new, high-mass particles, where the benefit from wavelet-based methods is assessed.

Finally, we summarise the thesis work, propose interesting lines of further research, and make a new cautious remarks regarding the use of wavelet analyses at current and future collider experiments.

I Theory

1.1 Particle Physics theory

The goal of the thesis is to develop techniques which will aid the search for new physics phenomena and their accompanying particles. But this requires that we have a sufficient understanding of the content and predictions of the currently accepted theories in particles physics. Therefore, this sections briefly presents the content of the Standard Model of particles physics and gives a minimal description of how to calculate probabilities for particle interactions. Finally, we mention some of the shortcomings of the Standard Model, and how these relate to diboson final states. The intention is not to present a complete account of the Standard Model, but rather to explain the essentials which will be needed later.

■ The Standard Model

The search for the fundamental constituents of the Universe is a corner stone in human intellectual endeavour, whose roots may be traced back to ancient philosophy, expressed in the term $\alpha\tau o\mu o\sigma$ (*atomos*, meaning 'indivisible'), coined by the Greek philosopher Democritus, ca. 400BC. This idea has survived to the present day where "fundamental" or "elementary" particles are those which are considered to have no constituents and no internal structure.

The pinnacle of this line of inquiry is the Standard Model (SM) of particle physics. The SM is a quantum field theory, which combines quantum mechanics and field theory to provide a unified description of a set of physical matter fields and their quanta—or *particles*. It is based on the principle of least action, which states that the trajectory followed by a system, through some configuration space, is that for which the action [1]

$$\mathcal{S} = \int \mathcal{L}(\phi, \partial_\mu \phi, x^\mu)\, \mathrm{d}^4 x \tag{1.1}$$

is stationary. Here $\mathcal{L}$ is the *Lagrangian* (density) for the system, ϕ is some field, and ∂_μ is the space-time derivative. The trajectory for which the action $\mathcal{S}$ is stationary can be shown to correspond to the solution to the Euler-Lagrange equation

$$\frac{\partial \mathcal{L}}{\partial \phi} - \frac{\partial}{\partial x^\mu}\left[\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)}\right] = 0 \tag{1.2}$$

which provides the equations of motions for the field ϕ. The Lagrangian therefore completely specifies the behaviour of the field, and can be written as

$$\mathcal{L} = \mathcal{L}_{\text{free}} + \mathcal{L}_{\text{int}} \tag{1.3}$$

where $\mathcal{L}_{\text{free}}$ describes the behaviour of the free field and $\mathcal{L}_{\text{int}}$ describes the interactions of the field (possibly, with other fields). The terms in the Lagrangian are not completely arbitrary, but must be invariant under a certain set of local symmetries; we say that they must be gauge invariant. In fact, the SM Lagrangian is the most general one, which is invariant under a particular set of symmetries, namely $SU(3) \times SU(2) \times U(1)$.

The symmetries of the SM naturally lead to the concept of gauge fields, representing forces, with associated field quanta: the force mediating particles, also called the *gauge bosons*. Therefore, the Standard Model is a theory of the fundamental particles of matter and their interactions.

Natural units

Before we proceed, we want to make a quick note in the idea of 'natural units'. In particle physics, it is customary to work in units, where $\hbar = c = 1$. For instance, this means that while energy is naturally given in units of electron-volts (eV $\approx 1.6 \times 10^{-19}$ J) or some derived unit (e.g. GeV $= 10^9$ eV), momentum and mass will have units of $[p] = [E]/c \rightarrow [E]$ and $[m] = [E]/c^2 \rightarrow [E]$. Therefore, all energetic quantities will be given in units of GeV—or similar—and proper SI units can always be retrieved by including factors of $\hbar$ and c where appropriate, based on dimensional analysis.

Particles and forces

The matter particles of the SM are fermions (half-integer spin particles) which come in two types: leptons and quarks. Each of these types form three distinct flavour doublets, collectively referred to as the three generations of matter. These are shown in Tab. 1.1.

Name	Flavour	Mass [MeV]	Charge [e]	Colour
Leptons				
Electron	e	0.5110		
Muon	μ	105.7	-1	
Tau	τ	1.777×10^3		
				0
Neutrinos				
Electron-neutrino	ν_e	$< 2.0 \quad \times 10^{-6}$		
Muon-neutrino	ν_μ	< 0.19	0	
Tau-neutrino	ν_τ	< 18.2		
Quarks (up-type)				
Up	u	2.3		
Charm	c	1.275×10^3	$+2/3$	
Top	s	$173.1 \quad \times 10^3$		
				R, G, B
Quarks (down-type)				
Down	d	4.8		
Strange	s	95.	$-1/3$	
Bottom	b	$4.18 \quad \times 10^3$		

Table 1.1: The fermions (all spin-1/2) of the Standard Model. All particles have anti-particles with opposite quantum numbers. Each line within each compartment correspond to a successively numbered generation (I, II, and III). Values from [2].

The leptons of the SM are the electron (e), muon (μ), and tau (τ) and their associated neutrinos (ν_e, ν_μ, and ν_τ), which form flavour

doublets, one for each generation:

$$\begin{pmatrix} e \\ \nu_e \end{pmatrix}, \quad \begin{pmatrix} \mu \\ \nu_\mu \end{pmatrix}, \quad \begin{pmatrix} \tau \\ \nu_\tau \end{pmatrix} \tag{1.4}$$

Generally, we will refer to e, μ, and τ collectively as 'leptons' (with symbol ℓ) and to ν_e, ν_μ, and ν_τ merely as 'neutrinos' (with symbol ν).

The quarks of the SM are the up (u), down (d), charm (c), strange (s), top (t), and bottom (b) quarks which similarly form flavour doublets, one for each generation:

$$\begin{pmatrix} u \\ d \end{pmatrix}, \quad \begin{pmatrix} c \\ s \end{pmatrix}, \quad \begin{pmatrix} t \\ b \end{pmatrix} \tag{1.5}$$

Generally, we will distinguish between up- (charge $+2/3$; u, c, and t, with symbol q_u) and down-type (charge $-1/3$; d, s, and b, with symbol q_d) quarks, collectively labeled q.

All particles in the SM have anti-particle which have the same mass and spin, but opposite quantum numbers (flavour, charge, colour, etc.; see below). These are marked with a bar: $\bar{f}$ is the anti-particle of f.

The interaction between matter particles is, as noted above, mediated by a set of gauge bosons. The bosons (integer spin particles) of the SM are shown in Tab. 1.2.

Table 1.2: The bosons of the Standard Model. Values from [2, 3]. Note, that the Higgs boson does not mediate a force.

Name	Label	Force	Mass [GeV]	Charge [e]	Colour	Spin
Gluon	g	Strong	0	0	8 states	
Photon	γ	EM	0	0		1
W boson	$W^\pm$	Weak	80.385	± 1	0	
Z boson	Z^0		91.188	0		
Higgs	H	—	125.09	0	0	0

Here, the gluons (g, of which there are eight) are responsible for the strong interaction, the photon (γ) for the electromagnetic force, and the massive bosons $W^\pm/Z^0$ (which we will refer to as the 'vector' or simply 'gauge' bosons) for the weak force. Finally, the Higgs bosons (H) is the only scalar (spin-0) particle in the SM and the only boson not mediating a force.

Forces in the SM are described as point-like interactions between fundamental particles, drawn as vertices in space-time with particles going in and out as "legs". A coupling is associated to each vertex, signifying the strength of that particular interaction (i.e. the probability for its occurrence). The interactions of the SM all satisfy some conservation laws, the most well-known being energy-momentum conservation. Below, we briefly describe the the different components of the SM, their interactions, and the associated conserved quantum numbers.

Electroweak theory

Electroweak (EWK) theory is the unified theory of the electromagnetic (quantum electrodynamics, QED) and weak interactions. The

electroweak force is mediated by four bosons, γ, $W^\pm$, and Z^0, and all EWK interactions are characterised by a general coupling strength[1] of [4] $\alpha \approx 1/137$.

Especially relevant to the EWK interactions is the conservation of electrical charge[2]. This is the quantum number to which the photon couples, with a strength proportional to the charge of the fermion, cf. the first vertex in Fig. 1.1. The conservation of electrical charge means, that the sum of charge flowing *into* a vertex must be exactly equal to the sum of charge flowing *out* of a vertex. Charge conservation is exact in all known interactions.
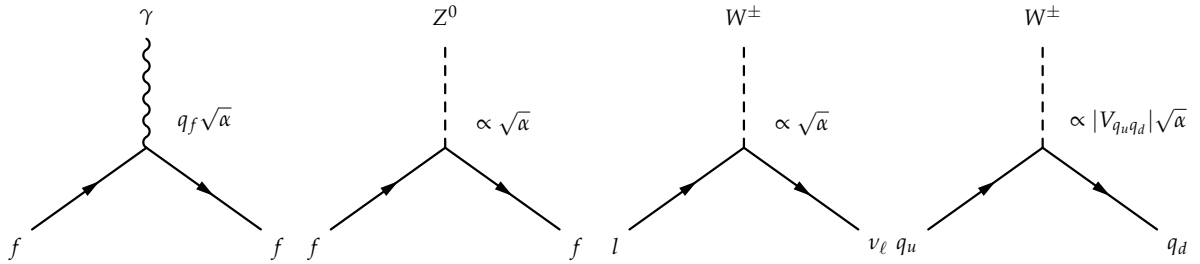
Additionally, as noted above, all fermions in the SM have a flavour which, for the leptons, corresponds to the matter generation. Each flavour has an associated quantum number called the lepton number L_ℓ, $\ell = e, \mu, \tau$ which is conserved in EWK interactions, and is defined as [5]

$$L_\ell \equiv N(\ell) - N(\bar{\ell}) + N(\nu_\ell) - N(\bar{\nu}_\ell) \tag{1.6}$$

where $N(\cdot)$ is the number of the particle in question in any interaction.

Similar quantum numbers exist for each quark species [5] (simply called strangeness, charm, etc.), which are also separately conserved in all interactions *except* the weak.

The neutral current (NC) interaction of the Z^0 boson conserves both charge and flavour quantum numbers. This means that its interaction with the fermions of the SM, cf. Tab. 1.1, will always be of the form $Z^0 f \bar{f}$, see the second vertex in Fig. 1.1, which is possible since it has no electrical charge. In addition, when coupling to charged fermions (i.e. all but the neutrinos), the interaction becomes similar to that for the photon, meaning that that the two processes will interfere, and in such cases they may collectively be referred to as[3] Z^0/γ^*.

On the other hand, the $W^\pm$ bosons, responsible for charged current (CC) interactions, has integer charge, meaning that interactions just with fermions and their anti-particles are not allowed by charge conservation. In the leptonic sector, CC interaction occurs exclusively within flavour doublets, see Eq. (1.4), on the form $W\ell\nu_\ell$, as shown by the third vertex in Fig. 1.1. Lepton universality means that the $W\ell\nu_\ell$ is the same for all lepton flavours. In the quark sector, since the quarks have fractional charges, charge conservation implies that CC interactions must involve an up-type and a down-type quarks, cf. Tab. 1.1, as shown in the fourth vertex in Fig. 1.1. Interestingly, the eigenstates of down-type quarks participating in charge current (CC) interaction are *not* the same as the flavour-eigenstates in Eq. (1.5).

[1] In this presentation we omit the discussion of the breaking of the $SU(2) \times U(1)$ structure of the electroweak sector, with associated couplings g_2 and g_1, and the corresponding mixing of the A_μ^a, with $a = 1, 2, 3$, and B_μ fields into the γ, $W^\pm$, and Z^0 bosons. The inherent difference in coupling strength between the photon and the massive gauge bosons is on the order of [2] $g_1/g_2 \tan\theta_W \sim \mathcal{O}(1)$, where θ_W is the Weinberg mixing angle.

[2] Given in units of the elementary charge e of the electron, see Tabs. 1.1 and 1.2.

Figure 1.1: Vertices for electroweak interaction and their order-of-magnitude couplings [2]. Here f is any fermion, q_u is an up-type quark, q_d is a down type quark, and $|V_{q_u q_d}|$ is the corresponding CKM matrix entry (see text for details). All vertices, here and in the following, may be drawn in a space-time diagram with any possible orientation. All diagrams throughout this thesis, unless explicitly stated otherwise, are made using the feynmf package, see https://www.ctan.org/pkg/feynmf.

[3] Where the asterisk indicates that the photon is off mass-shell, see below.

Instead, the $W^{\pm}$ bosons couple to the doublets

$$\begin{pmatrix} u \\ d' \end{pmatrix}, \quad \begin{pmatrix} c \\ s' \end{pmatrix}, \quad \begin{pmatrix} t \\ b' \end{pmatrix} \tag{1.7}$$

where d', s', b' are linear combinations of the down-type quarks (d, s, b) given by the unitary Cabbibo-Kobayashi-Maskawa (CKM) matrix

$$\begin{bmatrix} d' \\ s' \\ b' \end{bmatrix} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} \begin{bmatrix} d \\ s \\ b \end{bmatrix} \tag{1.8}$$

which may crudely be parametrised as [4]

$$|V_{q_u q_d}| = \begin{bmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{bmatrix} \approx \begin{bmatrix} 1 - \theta_C^2 & \theta_C & \theta_C^3 \\ \theta_C & 1 - \theta_C^2 & \theta_C^2 \\ \theta_C^3 & \theta_C^2 & 1 \end{bmatrix} \tag{1.9}$$

where $\theta_C \approx 0.2$ is the Cabbibo mixing angle. This shows that, although the majority of $W q_u q_d$ interaction are within the same generation, a non-negligible amount of inter-generational mixing does occur (e.g. Wus and Wcd). This means that flavour is strictly *not conserved* in charged current interactions, which are referred to as flavour-changing charged currents (FCCC).

Generally, for ease of notation, we will refer to the Z^0 boson just as Z and to the $W^{\pm}$ bosons collectively as W. Occasionally, we will refer to the massive gauge bosons collectively as $W/Z = V$.

Quantum chromodynamics

Quantum chromodynamics (QCD) is the quantum theory of the strong interaction. QCD interactions are mediated by the exchange of gluons and are characterised by the strong coupling constant α_S. Especially relevant to QCD is the colour quantum number, to which the gluons couple. This quantum number is conserved in all QCD interactions (which, it should be noted, is also the case for all EWK interactions), in addition to the charge and flavour quantum numbers.

The need for an additional quantum number became clear when studying the Δ^{++} state, composed of three u quarks in the 3/2-spin configurations [1]. Since quarks are fermions, the obey Fermi-Dirac statistics, which states that no two identical fermions may occupy the same state (i.e. have identical quantum numbers and overlapping wave-functions). The fact that the otherwise identical quarks can exist in the confined (uuu) state indicates the presence of an additional degree of freedom—or charge—which must have (at least) three distinct dimensions. The choice was made to call the charge 'colour' and label the three values 'red', 'green', and 'blue', or R, G, and B, with corresponding "anti-colours" $\bar{R}$, $\bar{G}$, and $\bar{B}$, similarly to the ∓ 1 electrical charges.

The quarks therefore carry one (anti-)colour charge to which the gluon couple. A major difference between QCD and QED, however, is that the gluons are *themselves* coloured[4], while photons carry no

[4] Such that each gluon has exist in a superposition of colour–anti-colour combinations, so as to ensure colour charge conservation are each vertex. There exist eight linearly independent such configurations, leading to a gluon multiplicity of eight, cf. Tab. 1.2.

electrical charge. Since the gluons are coloured themselves, and since gluons couple to the colour charge, the means that gluons exhibit self-interaction. The vertices of QCD are shown in Fig. 1.2.
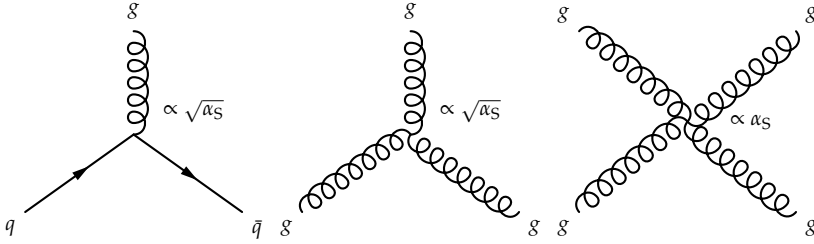

Figure 1.2: Vertices for strong interaction and their order-of-magnitude couplings [2].

This has the effect, that the strong coupling α_S isn't constant but depends on the distance scale or, equivalently, the momentum transfer Q^2, which is shown in the diagram in Fig. 1.3 for case of gluon-exchange between two quarks.

This behaviour is of the form [4]

$$\alpha_S(Q^2) \propto \frac{1}{\ln(Q^2/\Lambda_{QCD}^2)} \qquad (1.10)$$

with $\Lambda_{QCD} \approx 0.3\,\text{GeV}$. The strong coupling is said to be running[5]. It can then be measured at some fixed Q^2, e.g. $\alpha_S(m_Z^2) \approx 0.13$, and its behaviour with increasing Q^2 can then be found from Eq. (1.10).

Since distance scale and momentum transfer are inversely related by the Heisenberg uncertainty principle, α_S, and by implication the interaction between charged particles, diverges at large distances. This manifests itself in the quark–quark potential, which is of the form [5]

$$V(r) \approx -\frac{a}{r} + br \qquad (1.11)$$

and depicted in Fig. 1.4 for typical values of a and b. Here the r^{-1} term is the Coulomb potential and the r term is the colour potential.

The structure of this potential means that coloured states cannot exist in isolation beyond distances of $\mathcal{O}(\text{fm})$, leading to the concept of colour confinement. This means that all physically observable states must be colour neutral (i.e. have a net colour charge of "white"), which is why the colour charge was only postulated in 1964 [1].

We have already encountered a colour-neutral, physical state, namely the $\Delta^{++} = (uuu)$ particle introduced above. More generally, these colour-neutral states—called *hadrons*—may be split in to *baryons*—qqq states, like the proton $p = (uud)$ or neutron $n = (udd)$—and *mesons*—$q\bar{q}$ states, like the pions $\pi^0 = (u\bar{u})/(d\bar{d})$ and $\pi^{\pm} = (u\bar{d})/(\bar{u}d)$. Each hadron similarly has an anti-particle, with a content equal to the anti-particles of the original hadron, e.g. we have the anti-proton $\bar{p} = (\bar{u}\bar{u}\bar{d})$ and $\bar{\pi}^{\pm} = \pi^{\mp}$.

Interestingly, when we go the other way—towards larger Q^2-values or smaller distance scales—α_S decreases. This means that at very larger energies or very small distances (e.g. in the early Universe or inside hadrons, with radii of $\mathcal{O}(\text{fm})$) the strong coupling effectively
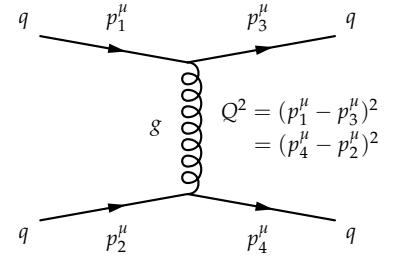
Figure 1.3: Gluon exchange by two quarks, and the associated momentum transfer Q^2.

[5] The same is true for α. However, since photons are electrically neutral they do not self-couple. Instead, the vacuum polarisation of photons sprouting virtual electron-positron pairs leads to a *screening* effect. This means that the EM coupling grows with Q^2 [4], but quite slowly, such that QED doesn't become strong until $Q^2 \sim \Lambda_{QED} > \Lambda_{GUT} \sim 10^{14}\,\text{GeV}$. The value of $\alpha \approx 1/137$ quoted earlier corresponds to $Q^2 = 0$. At $Q^2 = m_Z^2$ we have $\alpha(m_Z^2) \approx 1/129$.
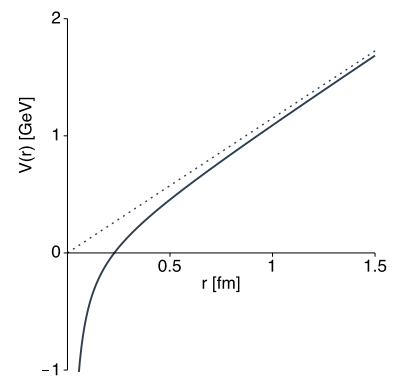
Figure 1.4: A model of the inter-quark potential.

vanishes, leaving quarks and gluons (collectively called "partons" inside hadrons) to propagate freely with virtually no interactions. This is called asymptotic freedom.

The Higgs field

The principle of gauge invariance means, that the mass terms in $\mathcal{L}_{\text{free}}$ are forbidden [1], implying that all particles in the SM should be massless. However, we know from experience, cf. Tabs. 1.1 and 1.2, that this is not the case. This can be remedied by introducing a new scalar[6] field ϕ, which was proposed independently by Brout and Englert [6], Higgs [7], and Guralnik, Hagen, and Kibbe [8] in 1964, and which is now just called the Higgs field. This field is postulated to have potential [1]

$$V(\phi) = \mu^2 |\phi|^2 + \lambda |\phi|^4 \tag{1.12}$$

A model of the potential is shown in Fig. 1.5.

If $\mu^2 < 0$ the lowest-energy state of the Higgs field is not at $\phi = 0$, but as some non-zero vacuum expectation value (vev) $\langle \phi \rangle = |\mu| / \sqrt{2\lambda}$. The Higgs field may then be parametrised as excitations around the vev, $\phi = \langle \phi \rangle + \phi_{\text{H}}$, where $\phi_{\text{H}}$ corresponds to the physical Higgs boson. The kinetic term for the Higgs field (which is gauge invariant) then results in ostensible mass terms for the W and Z bosons—but not for the photon, in compliance with experiment—as well as HVV interaction vertices with a coupling proportional to m_V^2. Similarly, we can introduce Yukawa terms to $\mathcal{L}_{\text{int}}$, relating the Higgs boson to the SM fermions, which in the same way leads to fermionic mass terms as well as Hff interactions with couplings proportional to m_f. The resulting vertices are shown in Fig. 1.6.

Although the exact form of the potential in Eq. (1.12) and well as the Higgs boson couplings and quantum numbers themselves are still mostly hypothetical, discoveries by the ATLAS [9] and CMS [10] experiments at the LHC are consistent with a SM Higgs boson with a mass of $m_H = 125.09\,\text{GeV}$ [3].

■ Perturbation theory

We now know the basic building blocks of particle interaction, given by the vertices in Figs. 1.1, 1.2, and 1.6. Ultimately, we are interested in the quantum mechanical amplitude for some initial configuration of n' particles, $|\alpha\rangle_{\text{in}}$ at time $t \to -\infty$, to evolve into some final configuration of m' particles, $|\beta\rangle_{\text{out}}$ at time $t \to \infty$, given some set of interaction rules. These transition amplitudes are given by the scattering matrix element, or simply matrix element, for some field ϕ [1]

$$\mathcal{M}_{\beta\alpha} = {}_{\text{out}}\langle \beta \,|\, \alpha \rangle_{\text{in}} \tag{1.13}$$

This amplitude is completely specified by the Green's function, or N-point correlation function, $G_N(x_1, \ldots, x_N)$, for $|\alpha\rangle_{\text{in}}$ and $|\beta\rangle_{\text{out}}$ having $n' + m' = N$ particles in total.
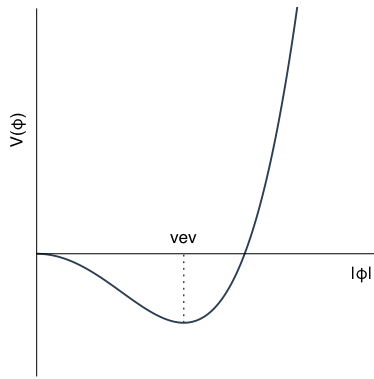
[6] Complex doublet.

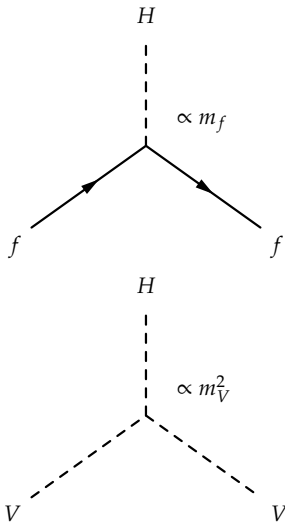Figure 1.5: A model of the Higgs potential with $\mu^2 < 0$.

Figure 1.6: Higgs boson vertices and their order-of-magnitude couplings [2]. Here f is any fermion and $V = W^\pm / Z^0$.

We are faced with one problem: any (connected) N-point graph may have an arbitrary number of internal vertices, which results in a divergent number of graphs, provided they all contribute equally.

To fix this, we note that the Green's functions for a field ϕ with a Lagrangian like Eq. (1.3) may be expressed in terms of the generating functional [1]

$$Z[J] = C \exp\left[i \int \mathrm{d}x\, \mathcal{L}_{\text{int}}\left(\frac{1}{i}\frac{\partial}{\partial J}\right)\right]$$
$$\times \exp\left[-\frac{i}{2}\int \mathrm{d}x\,\mathrm{d}y\, J(x)\Delta_{\text{F}}(x-y)J(y)\right] \qquad (1.14)$$

where C is a normalisation constant and $\Delta_{\text{F}}(x)$ is the Feynman propagator (in position space) for $\mathcal{L}_{\text{free}}(\phi)$. For a field with mass m and four-momentum p, Δ_{F} is generally on the form[7] (in momentum space)

$$\Delta_{\text{F}}(p) \propto \frac{1}{p^2 - m^2} \qquad (1.15)$$

But here the first factor describes the interaction (i.e. the vertices) of the theory—and only these—while the last factor just describes free particles propagating.

Assuming that $\mathcal{L}_{\text{int}}(\phi)$ contains vertices with some coupling g, and assuming furthermore that $g \ll 1$, the first factor in Eq. (1.14) may be written as a perturbative series in g, by writing put the expansion of $\exp(\,\cdot\,)$, where the n'th term counts all possible diagrams with $n-1$ vertices, with propagators between the vertices given by the expansion of the second factor. This mean that the transition amplitude (or matrix element) may be expressed entirely as a sum over diagrams, called Feynman diagrams, with weights given by (product of) the couplings of the internal vertices. Using the assumption that $g \ll 1$, we may truncate this perturbative series at some finite order in g (even just at lowest order, called 'tree level'), allowing us (in principle) to compute the full matrix element $\mathcal{M}_{\alpha\beta}$ up to this order. The probability for a certain process is then, like all quantum mechanical probabilities, found by squaring the sum of all relevant amplitude amplitudes, i.e. prob. $\propto |\sum \mathcal{M}_{\beta\alpha}|^2$. This which is the reason why the vertices in Figs. 1.1 and 1.2 carry weights proportional to the *square-root* of the coupling constants.

This can be generalised to the multiple fields (particles) of the SM, where perturbative series may be written in terms of α and α_{S}. One caveat is that, as noted above, $\alpha_{\text{S}} \to \mathcal{O}(1)$, in fact $\to \infty$, as $Q^2 \to \mathcal{O}(\Lambda_{\text{QCD}}^2)$ leading to a breakdown of the perturbation series. This mean that, while we are able to compute the amplitude for large-Q^2 QCD interactions, processes with $Q^2 \lesssim \mathcal{O}(\Lambda_{\text{QCD}}^2)$ are said to be non-perturbative and cannot be computed from theory.

▪ Vector bosons in searches for new physics

The Standard Model has proven extremely successful in describing and predicting the physical world, but it is, nevertheless, not complete. Several problems still plague the SM, since it doesn't describe [4]:

[7] Note here, that for internal propagators, in contrast to physical particles, we generally have $p^2 \neq m^2$ since they are virtual. We say that the propagator is off mass-shell. When $p^2 \to m^2$ we say that the propagator goes on mass-shell, and the amplitude blows up.

gravity; dark matter and dark energy; the vastly different mass-scales ($\Lambda_{\mathrm{QCD}}, \langle\phi\rangle, \Lambda_{\mathrm{Planck}} \sim 10^{19}$ GeV), leading to the hierarchy problem, which requires a fine-tuning of the Higgs mass of $\delta m_H^2 \sim \Lambda_{\mathrm{Planck}}^2$; and baryogenesis, i.e. the processes introducing the (observed) baryon asymmetry in the Universe, to name just a few.

For instance, it has been noted in [11] that the discovered Higgs mass of $m_H = 125.09$ GeV is too large to account for electroweak baryogenesis in itself, immediately requiring physics beyond the SM (BSM). As a possible solution, a minimal extension of the SM Higgs sector was proposed, by the addition of a massive scalar singlet. The allowed masses and mixing values are shown in Fig. 1.7.

Figure 1.7: Plot of the allowed scalar singlet masses and couplings between the Higgs states given current experimental limits and the condition of EWK baryogenesis.
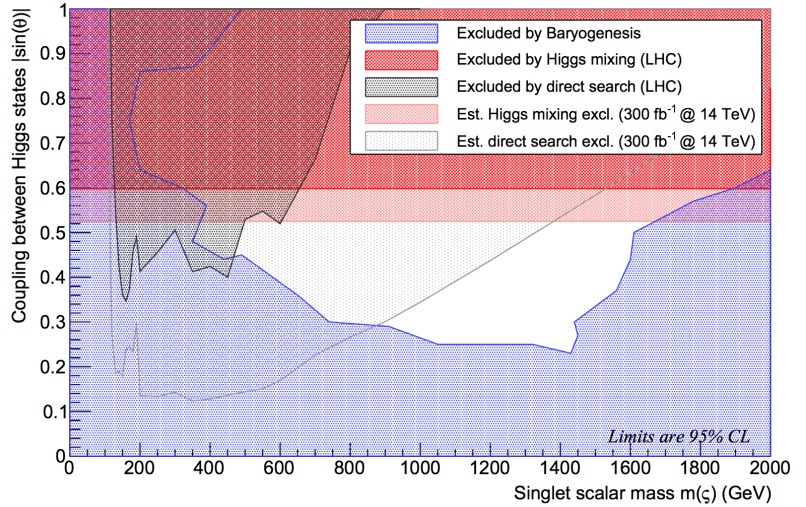
Although this is just one among a staggering amount of model for BSM physics, the allowed mass ranges are indicative of the regions which are relevant to current searches.

Diboson final states

Additional Higgs bosons may decay to final states with two vector bosons[8] through the second vertex in Fig. 1.6. In addition, several BSM models [12, 13] hypothesise new, narrow, high-mass resonances which similarly decay to diboson final states: $X \rightarrow W^{\pm}W^{\mp}/W^{\pm}Z^0/Z^0Z^0$. Current mass limits [14, 15, 16] are around $\mathcal{O}(800\,\mathrm{GeV})$ for neutral resonances (RS graviton) and $\mathcal{O}(1.5\,\mathrm{TeV})$ for singly charged ones (EGM W'), in agreement with the region of interest in Fig. 1.7.

This means that dibosons final states can be used as a model-agnostic channel, in which to search for signs of new physics. These diboson states subsequently decay either leptonically or hadronically (to quarks), and it is these decay products which are measured by experiments and those from which the decaying resonance is reconstructed. But which specific final state should one then focus on? To make this decision we need to know the intrinsic probability for the final state in question, as well as our ability to detect it.

[8] Actually, the $H \rightarrow ZZ^{(*)} \rightarrow 4\ell$ channel, with one resonant Z and one off-shell, was used in the discovery of the SM Higgs boson, despite $m_H < 2m_Z$ meaning that the process was Breit-Wigner suppressed, see Sec. 1.2.

The probability of a certain final state given by its branching ratio, which, for a certain process $R \to X$, is defined as

$$\text{BR}(R \to X) \equiv \frac{\Gamma(R \to X)}{\Gamma(R \to \text{anything})} \qquad (1.16)$$

where Γ is the width or inverse life-time. The branching ratio thus measures the fraction of decays of R which results in X.

The W bosons may decay to three distinct leptonic final states ($\ell\nu$), but the hadronic decays ($q\bar{q}$) are governed by the CKM-matrix, meaning that to first order only the diagonal elements contribute, such that only $W \to ud$, cs, and tb decays are allowed, where each is possible in three distinct colour states: $R\bar{R}$, $B\bar{B}$, and $G\bar{G}$. Additionally, since $m_t > m_{W/Z}$, the tb decay is Breit-Wigner suppressed, cf. Sec. 1.2, and is therefore negligible to first order. This leaves us with three $\ell\nu$ decays and two times three allowed $q\bar{q}$ decays, and since the $W\ell\nu$ and Wqq couplings are of the same order, cf. Fig. 1.1, we estimate

$$\text{BR}(W \to \ell\nu) \sim \frac{2}{3+6} = \frac{1}{3} = 33.\overline{3}\% \quad \to \quad 32.4\% \text{ (act.)} \qquad (1.17\text{a})$$

$$\text{BR}(W \to q\bar{q}) \sim \frac{6}{3+6} = \frac{2}{3} = 66.\overline{6}\% \quad \to \quad 67.6\% \text{ (act.)} \qquad (1.17\text{b})$$

which matches excellently with the actual, measured values [2].

The Z boson may decay both to $\ell\bar{\ell}$ and $\nu\bar{\nu}$, as well as to all same-flavour $q\bar{q}$ combinations except $t\bar{t}$ for the same reason as above. By an analogous counting to the above, we expect

$$\text{BR}(Z \to \ell\bar{\ell}) \sim \frac{3}{3+3+15} = \frac{1}{7} = 14\% \quad \to \quad 10.1\% \text{ (act.)} \qquad (1.18\text{a})$$

$$\text{BR}(Z \to \nu\bar{\nu}) \sim \frac{3}{3+3+15} = \frac{1}{7} = 14\% \quad \to \quad 20.0\% \text{ (act.)} \qquad (1.18\text{b})$$

$$\text{BR}(Z \to q\bar{q}) \sim \frac{15}{3+3+15} = \frac{5}{7} = 71\% \quad \to \quad 69.9\% \text{ (act.)} \qquad (1.18\text{c})$$

which is also in rough agreement[9] with the actual values [2].

We generally see that the branching to hadronic final states is large, but these are also very contaminated by backgrounds from SM processes, see Sec. 1.4. The leptonic modes have smaller branching ratios but are cleaner to detect experimentally (although the $Z \to \nu\bar{\nu}$ is invisible to detectors). A compromise, suitable for searches (which are limited by statistics) rather than precision measurements, are the semi-leptonic channels: $\ell\ell q\bar{q}$ and $\ell\nu q\bar{q}$. While the $\ell\ell q\bar{q}$ channels makes reconstructing the resonance invariant mass easier, cf. Sec. 1.4, we note that, assuming similar couplings of the RWW and RVV vertices, we expect

$$\frac{\text{BR}(R \to WW)}{\text{BR}(R \to ZZ)} = 2 \qquad (1.19)$$

since the W bosons are charged, while the Z boson is not, leading to a double counting. For this reason, we will use the neutral resonance $R \to WW \to \ell\nu q\bar{q}$ channel as our benchmark process for BSM resonance searches throughout this thesis, although the results will be applicable to the $\ell\ell qq$ and $qqqq$ channels as well.

[9] The remaining discrepancy comes from the differing weak isospin and charge quantum numbers for the various fermions, which factor into the Zff coupling.

1.2 Hadron collider physics

In the previous section we introduced the Standard Model, its particle content, and the rules for the interaction of these particles. We decided on a particle-level final state in which to look for new physics, and now turn to the question of how such high-mass resonances can be created through the collision of protons. We study the kinematics of their decay products and conclude with a detailed look at the anatomy of such processes as well as the way in which they may be studied through simulation.

■ Particle creation

As the Universe has cooled, all but the lightest particles have decayed, and become bound in hadrons and atoms. Therefore, in order to study high-mass particles, like the massive vector bosons W/Z, the Higgs boson H, or new high-mass particles not predicted by the Standard Model, we need a way to create them ourselves. This is done by colliding particles.

Initially, consider a $2 \to 1$ process, with two incoming particles interacting via the allowed vertices shown in Sec. 1.1 to form a third, outgoing particle. Let the two incoming particles have with same mass and opposite three-momentum, such that their for-momenta are $p_{1,2}^{\mu} = (E; \pm \mathbf{p})$. The invariant mass M of the two-particle state, and, therefore, of the outgoing particle, is then given by

$$M^2 = (p_1^{\mu} + p_2^{\mu})^2 = (2E)^2 - \mathbf{0}^2 = 4E^2; \qquad (p_1^{\mu} + p_2^{\mu})^2 \equiv \hat{s} \quad (1.20)$$

where $\hat{s}$ is the squared (hard scatter) center of mass energy. Therefore, in order to study particle with large masses M, we need to collide particles with increasingly large center of mass energies.

The next question is then which particles to collide. Choosing from the available stable particles, we have the fundamental choice of colliding leptons (i.e. not neutrinos) or hadrons. Leptons are, to our best knowledge, fundamental particles, the only stable one being the electron. These can can be accelerated to an energy of $M/2 \gg M_e$ each, and collided to produce new particles at exactly mass M. In addition to being able to fine-tune the hard-scatter center of mass energy, the choice of colliding leptons has the additional benefit of involving almost exclusively QED interaction, which is computable to all orders in perturbation theory and inherently free of the QCD-related complications regarding confinement and hadronisation, cf. Sec. 1.1. However, the fine-tuning also means that one has to choose the center of mass energy carefully, and *a priori* know exactly where to "look".

Hadrons, with protons as the prime example, on the other hand, are composite objects: quarks confined in colour-singlet states. For protons, the quantum numbers such as charge are satisfied by the (*uud*) 'valence' quarks, but each of these may radiate gluons, which in turn may sprout a "sea" of virtual quark/anti-quark pairs. This means that proton-proton (*pp*) collisions are much more involved than e.g. e^+e^- collisions, because it is not immediately clear exactly "what" we are colliding.

Parton density functions

For energies above the QCD confinement scale $\Lambda_{\text{QCD}} \sim \mathcal{O}(1\,\text{GeV})$, we can probe the internal, nuclear structure of protons (gluon and quarks), e.g. through deep inelastic scattering (DIS) [5]. Let $x = p/P$ be the fraction of the proton momentum P carried by any constituent parton. If the valence quarks alone made up the proton, these should all have $x = 1/3$. However, as is seen from the parton distribution functions (PDF) in Fig. 1.8, this is far from the case.

(a) $Q^2 = (2\,\text{GeV})^2$

(b) $Q^2 = (200\,\text{GeV})^2$
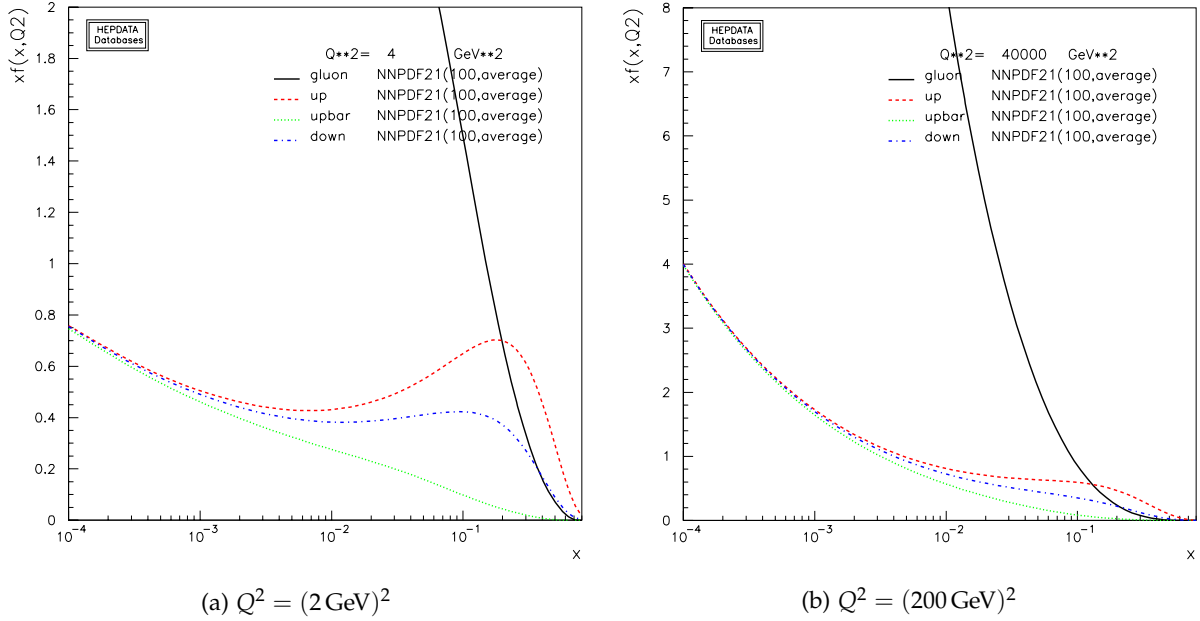
    The PDFs $f(x, Q^2)$ may be interpreted as the number density of partons carrying a momentum fraction x at a certain momentum transfer Q^2 [18]. The bumps of the u and d PDFs at $x \approx 1/3$ are clear proof of the presence of valence quarks, in rough proportion $2:1$ as expected. However, the $\bar{u}$ quark PDF shows that, particularly at very low x, sea quarks are numerous, and the gluon (g) PDF in particular makes it clear that the first-order (uud) view of the proton is insufficient. Furthermore, Fig. 1.8 shows that the "composition" of the proton changes with the momentum transfer: at lower Q^2, the valence quark structure is more pronounced, but as Q^2 increases, one is probing the proton with greater resolution, and is therefore sensitive to the non-valent partons.

    The PDF functions, both for gluons and quarks, can be parametrised as [19]

$$xf_i(x, Q^2) = A_i x^{a_i}(1 - x)^{b_i} \tag{1.21}$$

possibly with additional factors to allow for greater flexibility in fitting data. For instance, the high-x behaviour of the gluon PDF can [4] be approximated by $xf_g(x, Q^2) \propto (1 - x)^6$.

Figure 1.8: PDF's from the NNPDF 2.1 LO set [17], for u, $\bar{u}$, and d quarks as well as gluons. Plotted using the Durham HepData project PDF plotter, see `http://hepdata. cedar.ac.uk/pdf/pdf3.html`.

Factorisation theorem

We see that pp collisions are actually parton collisions, and we now have an understanding of what we are colliding (at different Q^2). Similarly, from Sec. 1.1, we have an understanding of how to compute probabilities for parton-level interactions, at least at energies sufficiently large that perturbation theory holds, which we now want to translate into a probability for creating a certain particle-level final state from incoming *protons*. This measure is called the cross section σ, which is defined as [5]

$$\sigma = \frac{\text{number of interaction per unit time per target}}{\text{incident flux}} \qquad (1.22)$$

The factorisation theorem now allows us to untangle the interactions in the pp collisions, factorising the contributions from the partonic interaction and the parton distribution information. For the production of any n-body final state X in a pp collision, the factorisation theorem reads [20]:

$$\sigma_X = \sum_{a,b} \iint_0^1 \mathrm{d}x_1\, \mathrm{d}x_2\, f_a(x_1; \mu_F^2)\, f_b(x_2; \mu_F^2) \quad \Big\} \text{ PDFs}$$

$$\times\, \hat{\sigma}_{ab \to X}\left(\{\hat{\mathbf{p}}_i^\mu\}; \alpha_S(\mu_R^2), \alpha(\mu_R^2), Q^2\right) \quad \Big\} \text{ hard process} \quad (1.23)$$

The second factor is the parton-level cross section for the $2 \to n$ process $ab \to X$, for any two partons allowing for such a process by the interactions specified in Sec. 1.1, where $\{\hat{\mathbf{p}}_i^\mu\}$ are the three-momenta of all $2 + n$ in-coming and out-going particles. The first factor then states, that the cross section for producing X from a pp collision is just the partonic cross sections, summed over all partons a, b and integrated over all possible configurations of momentum fractions $x_{1,2}$ with a weight given by the product of the corresponding PDFs. Here μ_F^2 is the factorisation scale and μ_R^2 is the renormalisation scale. These are usually set to $\mu_F^2 = \mu_R^2 = Q^2$.

The hard scatter cross section is, in turn, given by Fermi's Golden Rule for scattering [21]

$$\hat{\sigma}_{ab \to X} = \frac{\text{matrix element} \times \text{density of final states}}{\text{particle flux}}$$

$$= \underbrace{\frac{1}{2\hat{s}}}_{\text{Flux}} \int \underbrace{\prod_{i=1}^n \frac{\mathrm{d}^3 p_i}{2E_i (2\pi)^3}}_{n\text{-body phase-space}} \underbrace{(2\pi)^4 \delta^{(4)}\left(p_a + p_b - \sum_{i=1}^n p_i\right)}_{\text{Momentum conservation}} \underbrace{\overline{|M^{ab \to X}|^2}}_{\text{ME}}$$

$$(1.24)$$

Here the $1/\hat{s}$ term describes the parton flux [20], the Lorentz invariant phase space measures the density of final states (kinematics), and the matrix element, introduced in Sec. 1.1 and computed at some order in $\alpha_S(\mu_R^2)$, describes the dynamics of the process.

We now see the advantage that hadron collisions may have over lepton collisions: the experiments sets the hadronic center of mass

energy $\sqrt{s}$ but not the partonic one $\sqrt{\hat{s}} = \sqrt{x_1 x_1 s} \le \sqrt{s}$. This means that the pp collisions naturally scan a large range of $\sqrt{\hat{s}}$, weighted by the PDFs for the partonic constituents. Therefore, in cases where you *don't* know exactly what you are looking for, e.g. searches for new particles, hadron colliders provide a natural place to search broad mass ranges (since $M_X = \sqrt{\hat{s}}$), instead of performing pin-point measurements as $\ell^- \ell^+$ colliders would have you.

The downside of this, on the other hand, is that hadron collisions are inherently governed by QCD interactions, and therefore any hard partonic interaction breaks the colliding protons apart. Due to the confinement of colour charges, the proton remnants will undergo hadronisation, often resulting in considerable amounts of soft QCD activity ('underlying event') accompanying any hard scatter. This is a problem, which we will have to address.

■ Kinematics

An additional problem with colliding hadrons is that the center of mass system of the partonic interaction is will in general not coincide with the laboratory system. For instance, in a pp collision, the two interacting partons carry fractions $x_{1,2}$ of the respective proton momenta $\mathbf{P}_{1,2}$, with $\mathbf{P} \equiv \mathbf{P}_1 = -\mathbf{P}_2$. But then the parton center of mass system system will be moving with momentum $\mathbf{P}_{\text{cms}} = x_1 \mathbf{P}_1 + x_2 \mathbf{P}_2 = (x_1 - x_2)\mathbf{P}$. We therefore need to establish a set of coordinates which are invariant under Lorentz boosts along the direction of the incoming partons.

Rapidity and phase space

The usual set of coordinates is $p^\mu = (E, \mathbf{p}) = (E, p_x, p_y, p_z)$. If we align the z-axis with the beam direction $\mathbf{P}$, it is clear that the coordinates in the plane transverse to the beam direction, p_x and p_y, are unaffected by boosts β along z, while E and p_z are not, since the Lorentz transform takes the form [21]

$$\Lambda^\mu{}_\nu = \begin{pmatrix} \gamma & 0 & 0 & \gamma\beta \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \gamma\beta & 0 & 0 & \gamma \end{pmatrix}, \qquad p^{\mu\prime} = \Lambda^\mu{}_\nu p^\nu \qquad (1.25)$$

with Lorentz factor $\gamma = 1/\sqrt{1 - \beta^2}$, resulting in a transform which entangles E and p_z. Therefore, it makes sense to work in a cylindrical coordinate system with azimuthal angle ϕ (wrt. some x-axis) and transverse momentum $p_\perp \equiv \sqrt{p_x^2 + p_y^2}$, such that $p_x = p_\perp \cos\phi$ and $p_y = p_\perp \sin\phi$, both being boost-invariant. Additionally, the *invariant*[10] mass M, with $M^2 = p_\mu p^\mu = E^2 - |\mathbf{p}|^2$, of any state will also be a suitable coordinate. This means that we only need one more coordinate, specifying the z-component of the state.

We define a new variable, the rapidity y, as

$$y \equiv \frac{1}{2} \ln\left(\frac{E + p_z}{E - p_z}\right) = \ln\left(\frac{E + p_z}{M_\perp}\right) \qquad (1.26)$$

[10] Since the Lorentz transform by definition preserves distance elements [21] we have

$$g_{\mu\nu} \Lambda^\mu{}_\rho \Lambda^\nu{}_\sigma = g_{\rho\sigma}$$

which directly gives invariance of the mass

$$\begin{aligned} M'^2 &= p'_\mu p^{\mu\prime} = g_{\mu\nu} p^{\nu\prime} p^{\mu\prime} \\ &= g_{\mu\nu} \Lambda^\nu{}_\rho p^\rho \Lambda^\mu{}_\sigma p^\sigma = g_{\rho\sigma} p^\rho p^\sigma \\ &= p_\rho p^\sigma = M^2 \end{aligned}$$

[11] Since

$$E' = \gamma(E + \beta p_z)$$
$$p'_z = \gamma(p_z + \beta E)$$

we have

$$
\begin{aligned}
y' &= \frac{1}{2}\ln\left(\frac{E' + p'_z}{E' - p'_z}\right) \\
&= \frac{1}{2}\ln\left(\frac{\gamma(E + \beta p_z) + \gamma(p_z + \beta E)}{\gamma(E + \beta p_z) - \gamma(p_z + \beta E)}\right) \\
&= \frac{1}{2}\ln\left(\frac{1+\beta}{1-\beta}\frac{E + p_z}{E - p_z}\right) \\
&= \frac{1}{2}\ln\left(\frac{E + p_z}{E - p_z}\right) + \underbrace{\ln\left(\sqrt{\frac{1+\beta}{1-\beta}}\right)}_{\equiv k} \\
&= y + k
\end{aligned}
$$

which proves that rapidity is additive under boosts along the z-axis.

[12] We can express the equivalence of the two sets of coordinates as

$$
\begin{aligned}
p^\mu &= (E; p_x, p_y, p_z) \\
&= (M_\perp \cosh y; \\
&\qquad p_\perp \cos\phi, p_\perp \sin\phi, M_\perp \sinh y)
\end{aligned}
$$

The x- and y-components follow directly from the definition of $p_\perp$ and the choice of (η, ϕ, z) coordinate system. The E- and z- components follow from

$$
\begin{aligned}
M_\perp^2 &\equiv E^2 - p_z^2 \\
&= M_\perp^2 \cosh^2 y - M_\perp^2 \sinh^2 y \\
&= M_\perp^2 \underbrace{(\cosh^2 y - \sinh^2 y)}_{1} = M_\perp^2
\end{aligned}
$$

and

$$
\begin{aligned}
y &\equiv \frac{1}{2}\ln\frac{E + p_z}{E - p_z} \\
&= \frac{1}{2}\ln\frac{M_\perp \cosh y + M_\perp \sinh y}{M_\perp \cosh y - M_\perp \sinh y} \\
&= \frac{1}{2}\ln\frac{(\cosh y + \sinh y)^2}{\underbrace{\cosh^2 y - \sinh^2 y}_{1}} \\
&= \ln(\cosh y + \sinh y) \\
&= \ln\left(\frac{1}{2}\left[(e^y + e^{-y}) + (e^y - e^{-y})\right]\right) \\
&= \ln e^y = y
\end{aligned}
$$

whereby we establish the second equality in Eq. (A.2).

with transverse mass [19] $M_\perp = \sqrt{M^2 + p_\perp^2} = \sqrt{E^2 - p_z^2}$. This variable is additive under boosts[11] along the z-axis, meaning that *differences* in y are Lorentz invariant invariant. Therefore, in rest of this study, we will mainly be working with the coordinates M, $p_\perp$, ϕ, and y, which are equivalent[12] to the standard set choice of $(E; p_x, p_y, p_z)$.

In these coordinates, the Lorentz invariant phase-space element [21] becomes:

$$d\Phi = \frac{dp_x\, dp_y\, dp_z}{E} \to p_\perp\, dp_\perp\, d\phi\, dy \tag{1.27}$$

which yields a phase space uniform in ϕ and y. This will be of importance later, for the discussion of particle detectors in Sec. 1.3.

We notice that, for a combined two-parton system, we have

$$E = (x_1 + x_2)|\mathbf{P}| = (x_1 + x_2)\frac{\sqrt{s}}{2} \tag{1.28a}$$

$$p_z = (x_1 - x_2)|\mathbf{P}| = (x_1 - x_2)\frac{\sqrt{s}}{2} \tag{1.28b}$$

so that the system momentum transfer, or invariant mass, is

$$
\begin{aligned}
Q^2 = M^2 = E^2 - p_z^2 &= \left[(x_1 + x_2)^2 - (x_1 - x_2)^2\right]\left(\frac{\sqrt{s}}{2}\right)^2 \\
&= x_1 x_2 s \tag{1.29}
\end{aligned}
$$

and the system rapidity can be written as, cf. Eq. (1.26)

$$y = \frac{1}{2}\ln\frac{x_1}{x_2} \tag{1.30}$$

We have now linked a quantity at the parton scale (x) with one we can measure $(y$ of the final state system), meaning that given a measured system y we can relate the two by [4]

$$x_{1,2} = \frac{M}{\sqrt{s}}e^{\pm y} \tag{1.31}$$

But since, for any given system invariant mass M we must have $x_{1,2} \leq 1$, the maximal possible rapidity is

$$|y| \leq y_{\max} = \ln\frac{\sqrt{s}}{M} \tag{1.32}$$

Therefore, the decay products are confined to an allowed range of system system rapidities of $\Delta y = 2y_{\max}$. In addition, we just saw that phase space is roughly uniform in y. This gives rise to the concept of a "rapidity-plateau", where final state products are distributed in a roughly uniform manner on $|y| \lesssim y_{\max}$.

Pseudo-rapidity and distance measures

We now have two coordinates which are purely energetic $(m$ and $p_\perp)$ and two which have directional information $(\phi$ and $y)$. This allows us to define a Lorentz invariant distance measure in our cylindrical coordinate system

$$dR^2 = dy^2 + d\phi^2 \tag{1.33}$$

where dy and $d\phi$ are (not necessarily infinitesimal) differences in the respective coordinates. There is one caveat about the notion of rapidity, however: it is mass dependent and therefore the distance measure in Eq. (1.33) is not completely geometric. This is not a problem as such, and will be used later when discussing jets in Sec. 1.4. However, if we are studying massless particles, or if we are unable to measure particle masses and set $m = 0$, the expression in Eq. (1.26) simplifies as

$$y \xrightarrow[m=0]{} \eta \equiv \frac{1}{2} \ln \left(\frac{|\mathbf{p}| + p_z}{|\mathbf{p}| - p_z} \right) = \frac{1}{2} \ln \left(\frac{|\mathbf{p}| + |\mathbf{p}|\cos\theta}{|\mathbf{p}| - |\mathbf{p}|\cos\theta} \right)$$

$$= \frac{1}{2} \ln \left(\frac{1 + \cos\theta}{1 - \cos\theta} \right) = -\ln \tan \frac{\theta}{2} \qquad (1.34)$$

where θ is the angle of $\mathbf{p}$ wrt. the z-axis. The quantity η is called the pseudo-rapidity, and is identical to y for massless particles. Differences in pseudorapidity for massive particles are *not* invariant under Lorentz transform, but on the other hand Eq. (1.34) defines a purely geometric quantity, which can serve as the longitudinal coordinate in our cylindrical system. We can define a distance measure for pseudorapidity similar to that in Eq. (1.33)

$$dR^2 = d\eta^2 + d\phi^2 \qquad (1.35)$$

Due to their similarity we will refer to Eqs. (1.33) and (1.35) interchangeably. But, due to its relative simplicity and purely geometric nature, we will often work in coordinates using pseudorapidity, rather than rapidity, despite it not being completely Lorentz invariant.

Resonances

AS already mentioned, this study is motivated by the search for particles not predicted by the Standard Model, which will manifest themselves as resonances R with mass M. These will be produced in s-channel processes like the one shown in Fig. 1.9.

s-channel production at tree level is known [21] to have a matrix elements which behave as the propagator of the intermediate particle, cf. Eq. (1.15)

$$\mathcal{M} \sim \frac{1}{\hat{s} - M^2} \qquad (1.36)$$

Furthermore, resonances not predicted by the Standard Model are likely to have high masses, and therefore be unstable. If the resonance has a width of Γ (i.e. a lifetime of $\tau = 1/\Gamma$) the propagator is further modified as

$$\mathcal{M} \sim \frac{1}{\hat{s} - M^2 + iM\Gamma} \quad \rightarrow \quad |\mathcal{M}|^2 \sim \frac{1}{(\hat{s} - M^2)^2 + M^2\Gamma^2} \qquad (1.37)$$

Since $\sqrt{\hat{s}}$ is the invariant mass of any hard interaction, we conclude that a new, heavy particle will show up as an enhancement in the invariant mass distribution for a particular final state, the shape of the resonance given by the Breit-Wigner distribution in Eq. (1.37).
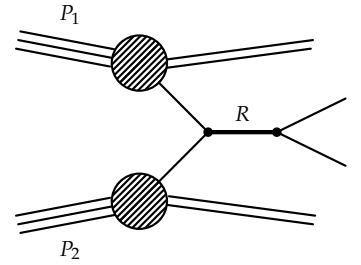

Figure 1.9: Resonance production in the s-channel. The three lines going into the large "blobs" represent the three valence quarks in each proton $P_{1,2}$. The large blobs indicate the factoring out of the hard scatter partons, interacting to create the resonance, from the protons. The two times two out-going lines (top and bottom) are the proton remnants.
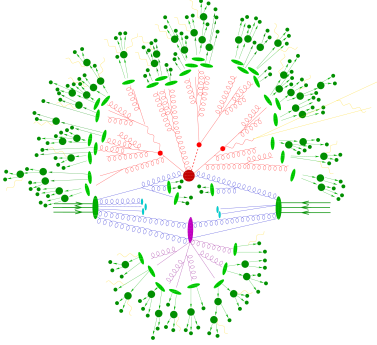
Figure 1.10: A schematic representation of incoming (anti-)protons incoming (anti-)protons (green lines), making up the initial state partons incl. radiation (blue), which collide in a hard interaction (large red blob), with subsequent heavy particle decay (smaller red blobs), as well as a second partonic interaction (purple blob), both with subsequent final state radiation (red and purple), and hadronisation (green blobs). See text for details. From [22].

Event generators and the anatomy of a hadron collision

In the search for new physics beyond the Standard Model, we are looking for tiny variations—like the resonance-bumps above—amidst an overwhelming number of regular, SM processes. Therefore, if we are to search for such minuscule variations, we need precise predictions for the expected SM processes; not just in terms of total cross sections for various processes, but also their kinematic distributions. However, as we have alluded to already, this is an involved task in the setting of hadron collision. Diagrams such as Fig. 1.9 are deceptively simplified, and do not do justice to the complicated structure of a hadron collision: we are colliding partons, initially in a colour-confined state; the processes are inherently non-perturbative at low energies; and all partons in the final state must ultimately recombine into colour-singlet states. A more detailed representation is given in Fig. 1.10.

Monte Carlo (MC) techniques, i.e. ones relying on random sampling to generate simulated events, allows for a detailed representation of hadron collisions, which would be impossible to study analytically. Several general purpose MC event generators, among these PYTHIA [23], HERWIG [24], and SHERPA [22], along with numerous more specialised programs. These all utilise the factorisation of different components of hadron collisions, as exemplified by Eq. (1.23), allowing for the compartmentalisation of the MC generators into various specialised sub-routines. In the rest of this section, we will describe the anatomy of a hadron collision and, to the extent relevant for this study, the implementation in the Monte Carlo programmes used to simulate such collisions.

Hard scatter matrix element

The central component of any (hard) pp collision is the hard scatter, marked by a large red blob in Fig. 1.10. The probability for any given hard scatter process is given by its matrix element (ME) as evident from Eq. (1.24). As noted before, this is the squared sum of all Feynman graphs contributing to a particular process, representing a perturbative series in α_S to some finite order. For a certain final state, e.g. the production of a W boson, the leading order (LO) calculation is at order $\mathcal{O}(\alpha)$ and given by the diagram in Fig. 1.11(a); see also the fourth vertex in Fig. 1.1. However, these processes only constitute the lowest order in the perturbation series, and with the need for transverse activity (see Sec. 3.1) as well as the ever increasing precision of modern particle experiments, it is necessary to go to higher orders.

Figs. 1.11(b) and 1.11(c) show two of the diagrams contributing to the $q\bar{q} \rightarrow W$ ME at next-to-leading order (NLO). However, using just diagrams of these types, if we want to compute observables involving the W boson and the additional partons, the results will *not* be correct at NLO, only at LO [18]. Instead, the diagram in Fig. 1.11(c) contributes to the LO ME for the $q\bar{q} \rightarrow W + 1$ additional parton "leg" process. Similarly, Fig. 1.11(d) constitutes a contribution to the

NNLO $q\bar{q} \to W$ ME *or* to the NLO $q\bar{q} \to W + 1$ leg ME *or* to the LO $q\bar{q} \to W + 2$ legs ME. The full ME at N^{n}LO is calculated as the sum over all diagrams with $k + l \leq n$, where k is the number of additional legs and l is the number of additional loops [18].

Parton showering

While the computation of LO matrix elements has been automated, higher order diagrams are increasingly difficult to compute [25]. Therefore MC generators employ parton showering (PS) to approximate the effect of additional legs to all higher orders [20].

Consider again the production of a W, now with subsequent hadronic decay $W \to q\bar{q}$. We can compute the ME for the same process with an additional final state gluon, radiated from either of the quarks, i.e. $W \to q\bar{q}g$, and it may be parametrised as [20]

$$\frac{\mathrm{d}\sigma_{q\bar{q}g}}{\sigma_{q\bar{q}}} \approx \sum_{\text{parton}} \frac{\alpha_{\mathrm{S}}}{2\pi} \frac{\mathrm{d}\theta^2}{\theta^2} \, \mathrm{d}z \, \frac{1 + (1-z)^2}{z} \qquad (1.38)$$

Here z is the fraction of the radiating quark's momentum carried away by the gluon, and θ is the opening angle between the radiated gluon and the emitting quark. We see that the $q\bar{q}g$ cross section is proportional to the $q\bar{q}$ cross section, which means that the right hand side of Eq. (1.38) can be seen as the contribution from the $q \to qg$ splitting, factoring out. The contribution is seen to diverge in the the collinear ($\theta \to 0$) and soft gluon ($z \to 0$) limit, highlighting the importance of considering the effects of additional parton radiation beyond the ME. This process is not specific to the $q \to qg$ splitting, and the factorisation observed in Eq. (1.38) also holds for $g \to gg$ and $g \to q\bar{q}$, meaning that we can generalise the emission of additional partons as [20]

$$\frac{\mathrm{d}\sigma_{bcd}}{\sigma_{ad}} \approx \sum_{\text{parton } a} \mathrm{d}\mathcal{P}_{a \to bc}, \qquad \mathrm{d}\mathcal{P}_{a \to bc} = \frac{\alpha_{\mathrm{S}}}{2\pi} \frac{\mathrm{d}\theta^2}{\theta^2} P_{a \to bc}(z) \, \mathrm{d}z \quad (1.39)$$

where $P_{a \to bc}(z)$ are the DGLAP splitting kernels [26]

$$P_{q \to qg}(z) = \frac{1 + (1-z)^2}{z} \qquad (1.40\mathrm{a})$$

$$P_{\bar{q} \to \bar{q}g}(z) = P_{q \to qg}(z) \qquad (1.40\mathrm{b})$$

$$P_{g \to gg}(z) = \frac{z^4 + 1 + (1-z)^4}{z(1-z)} \qquad (1.40\mathrm{c})$$

$$P_{g \to q\bar{q}}(z) = z^2 + (1-z)^2 \qquad (1.40\mathrm{d})$$

up to factors of order unity. These are universal, which means that they allow for a process-independent implementation in MC event generators. Also, these branchings occur not only for final state partons, but also for incoming (initial state partons): these continuously emit and recombine virtual gluons and $q\bar{q}$-pairs, on a time scale related to their energy by the Heisenberg uncertainty principle [27]. However, the virtual partons created immediately before the hard
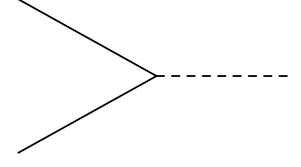
(a) LO $q\bar{q} \to W$.
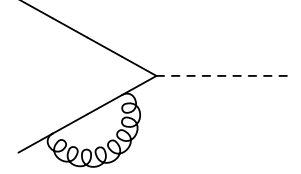
(b) NLO $q\bar{q} \to W$.
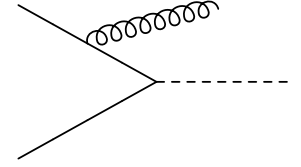
(c) NLO $q\bar{q} \to W$ *or*
LO $q\bar{q} \to W + 1$ leg.
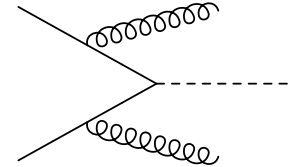
(d) NNLO $q\bar{q} \to W$ *or*
NLO $q\bar{q} \to W + 1$ leg *or*
LO $q\bar{q} \to W + 2$ legs.

Figure 1.11: Diagrams contributing to $W + n$ legs matrix elements at various orders.

interaction do not have time to recombine and thus manifest themselves as radiation from the incoming partons. These two categories of parton radiation are called final and initial state radiation (FSR and ISR) and are marked in green and blue, resp., in Fig. 1.10. These processes are handled separately in MC generators.

Finally, we note that instead of the opening angle θ, the splitting probabilities in Eq. (1.39) may instead be parameterised [20] by the radiated (off-shell) parton's virtuality or mass $Q^2 = z(1-z)\theta^2 E^2 = M^2$ (where E is its energy) or by its momentum transverse to the emitting parton $p_\perp^2 = z^2(1-z)^2\theta^2 E^2$ since

$$\frac{d\theta^2}{\theta^2} = \frac{dQ^2}{Q^2} = \frac{dM^2}{M^2} = \frac{dp_\perp^2}{p_\perp^2} \tag{1.41}$$

We see that the two variables parametrising the parton radiation, z and either of the ones in Eq. (1.41), are different in character [27]: the virtuality variable Q^2 is "temporal" in that it has memory of the preceding evolution, since the virtuality of the emitting parton decreases for each splitting, while z is purely "spatial" (has no memory of the preceding evolution). Therefore, the PS procedure starts from the final state partons from the ME which, for hard collisions, will have large energies. We can then perform an iterative evolution down in some temporal evolution variable (e.g. an angular ordering), resulting in a shower of radiated partons[13], until all final state particles reach energies of $\mathcal{O}(\Lambda_{QCD})$.

Hadronisation

At energies of $\mathcal{O}(\Lambda_{QCD})$, QCD becomes non-perturbative and the parton showering process breaks down. Since we know from Sec. 1.1 that all particles must exist in colour-singlet states, MC event generators need prescriptions to enforce such a colour-neutral recombination. As described in [18], the hadronisation process is simulated through three steps:

1. partons are grouped into a continuum of high-mass hadronic states, either through colour "strings" (PYTHIA) or "clusters" (HERWIG);

2. these hadronic states are then iteratively decomposed into a collection of primary hadrons through string fragmentation or cluster splitting;

3. the non-stable primary hadrons are then decayed into stable products.

Using the Lund string model [20] as an example, the final state partons resulting from the ME plus ISR and FSR are connected by colour strings, ensuring the conservation of colour charge. These string may have complicated topologies, with serveral "kinks" connecting partons from the same decay chain. The central idea is then to use the concept of linear colour confinement at large distances, cf. Eq. (1.11). As partons move apart, the potential energy in the string increases,

[13] We note, that the divergences in Eqs. (1.39) and (1.40) will in principle result in the emission of infinitely many soft/colliear partons. This has no consequences for physical measurements since the the number of final state partons above some finite energy threshold is always finite. These divergences are avoided in MC generators, e.g. by simple lower limits on z and Q^2.

eventually allowing virtual $q\bar{q}$ pairs to become real by breaking the string, see Fig. 1.12.

This process proceeds, until all partons are recombined into colour-neutral states with insufficient energy to cause further string fragmentation. The remaining colour singlet state are then identified as hadrons.

Figure 1.12: Two partons moving apart, connected by a Lund colour string building up potential energy, eventually breaking to form a new $q\bar{q}$ pair. From [18].

Soft physics and pile-up

Apart from the hard interaction and associated radiation, the pp collision in Fig. 1.10 contains other (soft) activity which is collectively labeled the underlying event (UE). This covers the remnant of the non-interacting parts of the colliding protons as well as the possibility of multiple parton interactions (MPI) within the same pp collision as the hard scatter, Fig. 1.10. This soft physics interferes with the hard scatter and acts to "obscure" the details of the hard interaction, which is the process in which we are interested. However, even in the absence of a hard interaction, soft dissociative interactions occur. In fact, these soft events (called minimum bias, minbias, or MB) are by far the most common type of events observed at hadron colliders, making up[14] around 2/3 of the total cross section [27]. To leading order, these processes are dominated by soft $2 \rightarrow 2$ parton interactions. For tree-level parton exchange in the t-channel, see e.g. Fig. 1.3, the partonic cross section goes as [18]

$$\frac{\mathrm{d}\sigma_{2\rightarrow 2}}{\mathrm{d}\hat{p}_T^2} \approx \frac{1}{\hat{p}_\perp^4} \tag{1.42}$$

Minbias events have no real structure, and mostly just manifest themselves as a diffuse "noise" of soft particles, roughly uniform in y and ϕ, cf. Eq. (1.27). Previous measurements [28] have found that, requiring a particle $p_\perp > 500\,\mathrm{MeV}$, each minbias collision at $\sqrt{s} = 7\,\mathrm{TeV}$ will result in a roughly 3 charged particles per unit of pseudorapidity, each carrying on average $\langle p_\perp \rangle \approx 0.8 - 1\,\mathrm{GeV}$. Even within the same bunch crossing as a particular hard interaction, several of these soft minbias interactions will occur, resulting in a considerable number of final state particles, affecting the measurement of the hard interaction. The sum of these simultaneous soft collisions are called *pile-up*, and are discussed further in Sec. 2.2.
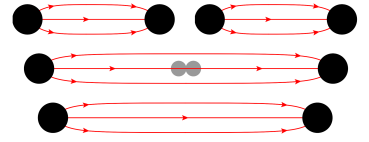
[14] The concept of minimum bias events is only meaningful provided a specification of the triggering used to *define* what is meant by "minimum". However, since we treat minbias events only as "soft physics" events, without resort to their detailed definition, the discussion of such trigger specifications is not of primary interest in the present study.

1.3   The ATLAS detector

In the previous section we described in some detail the creation of particles in hadron collisions, the kinematics and decays of the outgoing particles as well as the process of hadronisation. The question is now how to measure the final state particles emerging from these collisions. Therefore, this section gives a short overview of the ATLAS experiment at the LHC, describing the operating principles, purpose, design, and performance of each of the different sub-detectors, to the extent relevant for this study.

■   The Large Hadron Collider

[15] European Organization for Nuclear Research ("Conseil Européen pour la Recherche Nucléaire").

The CERN[15] accelerator complex was founded in 1954, with the goal of studying the smallest constituents of matter: atoms, fundamental particles, and their interaction. Here, the largest particle accelerator ring in the world, the LEP ring with a circumference of 26.7 km, was finished in 1988 and now houses the Large Hadron Collider (LHC). From the onset of its operation in 2008, the LHC represents the state of the art of man-made particle acceleration, performing proton-proton collisions at center of mass energies of $\sqrt{s} = 7$ and 8 TeV during its first period of operation (Run 1, 2009–2013) with nominal energies of $\sqrt{s} = 13$ and 14 TeV expected during the second (Run 2, 2015–2018).

Particle acceleration

The LHC [29], a super-conduction accelerator and storage ring with two counter-rotating proton beams, is the last step in an injection chain, gradually accelerating protons from rest to beam energies of up to 7 TeV. The first step is the Linear Accelerator 2 (Linac2), which uses an array of three radio-frequency (R-F) accelerator tanks [30], with a combined length of 33.3 m, to accelerate protons from $\sim 0 - 50$ MeV. The beam is then fed into a series of three proton synchrotrons—Proton Synchrotron Booster (PSB, circumf. = 157 m); Proton Synchrotron (PS, circumf. = 628 m); Super Proton Synchrotron (SPS, circumf. = 7 km)—each increasing the beam energy to 1.4, 25, and 450 GeV resp., using R-F cavities along the individual ring circumferences to increase the proton energies, and synchronously increasing dipole fields to steer the beams. In order to ensure sufficiently long beam lifetimes, the protons are accelerated and collided in vacuum pipes with very stringent requirements: in the LHC beam pipe the gas density is required to be below a equivalent hydrogen density of $10^{15}\,H_2\,m^{-3}$ [29].

Luminosity

Having reached its target beam energies, the LHC collides then two counter-rotating proton beams at four interaction points, each holding one of the major LHC experiments: ATLAS, CMS, LHCb, and ALICE. Nominally, each beam holds 2808 proton bunches, each containing $\mathcal{O}(10^{11})$ protons per bunch [29], with a 25 ns bunch spacing. During

Run 1, the number of bunches in each beam was chosen to be half of the nominal.

Given a cross section σ for some process, the total interaction rate dN/dt and, correspondingly, the total number of collisions—or "events"—N of this type, is given by

$$\frac{dN}{dt} = \mathcal{L} \times \sigma \qquad \longrightarrow \qquad N = \int \mathcal{L} \, dt \, \sigma \qquad (1.43)$$

where $\mathcal{L}$ is the instantaneous machine luminosity, which is a measure for the pp flux at each bunch crossing. The nominal peak instantaneous luminosity for the LHC is $\mathcal{L} = 10^{34} \, \mathrm{cm}^{-2} \, \mathrm{s}^{-1}$ [29]. The instantaneous luminosity integrated over some run duration, the integrated luminosity $\int \mathcal{L} \, dt$, measured in units of inverse barn ($1 \, \mathrm{b} = 10^{-24} \, \mathrm{cm}^2$), is a measure of the total amount of data collected, and we will frequently refer to this quantity simply as 'luminosity' with the symbol $\mathcal{L}$, relying on the context and choice of units to distinguish the integrated and instantaneous luminosity.

The LHC beam lifetime is around $\tau_L = 15$ hours and is mainly determined by beam collisions. With such a lifetime, the optimal run duration is between 6 and 12 hours, in principle allowing for a total integrated luminosity collection of[16] $\mathcal{L} \approx 100 \, \mathrm{fb}^{-1}/\mathrm{year}$ [29]. However, during the LHC Run 1 a total of $\mathcal{L} \approx 30 \, \mathrm{fb}^{-1}$ was delivered by the LHC, for the $\sqrt{s} = 7$ and $8 \, \mathrm{TeV}$ runs combined .

The total cross section at the LHC is roughly $\sigma_{\mathrm{tot.}} = 100 \, \mathrm{mb}$ [31] which means that, without performing any selection of presumed interesting final state, we expect an interaction rate of a staggering $dN/dt \approx 10^9$ events/second at the nominal instantaneous luminosity. This clearly shows that we need to design experiments which will be able to discard this vast number of uninteresting events, while storing the few events which might provide hints regarding BSM physics.

[16] $1 \, \mathrm{fb}^{-1} = 1 \, \mathrm{femtobarn}^{-1} = 10^{15} \, \mathrm{b}^{-1}$.

The ATLAS experiment at the LHC

One of the experiments doing just this is the ATLAS detector [32], which, along with CMS [33], is one of the two LHC general-purpose detectors, designed for precision measurements of the Standard Model as well as observation of new phenomena at the TeV-scale.

ATLAS, as shown in Fig. 1.13, has a cylindrical layout, for the reasons discussed in Sec. 1.2, and is described in a coordinate system with origin in the nominal interaction point (IP) at the center of the detector, where the pp collisions occur. The detector is forward-backward symmetric in the plane transverse to the beam direction, with the z-axis pointing in the counter-clockwise direction around the LHC ring, and the x-axis pointing towards its center. The ATLAS sub-detectors are generally segmented in bins of η and ϕ, since we saw in Sec. 1.2 that the production of (soft) particles is expected to be roughly uniform in y and ϕ, and using the pseudorapidity approximation allows for a purely geometric subdivision. Starting from the IP, the final state particles will traverse the detector on their outwards path, passing first the inner detector (ID), the electromagnetic and
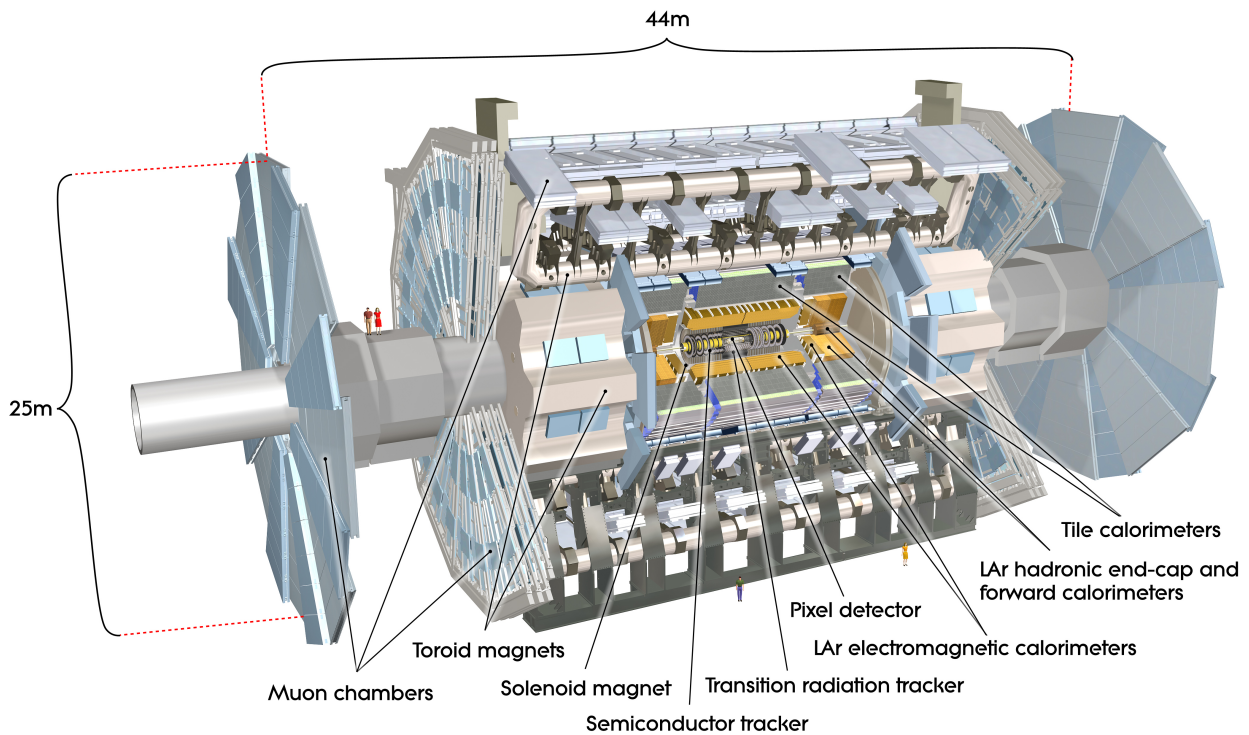
Figure 1.13:   Overview of the
ATLAS detector.   Image from
[CERN-GE-0803012], ATLAS Exper-
iment © 2015 CERN.

[17] Giving a B-field parallel to the z-axis,
bending the charged particles in the
transverse plane.

[18] Producing an azimuthal magnetic
field, being charged particles in the
$R - z$ plane.

hadronic calorimeters (ECAL and HCAL, resp.), and finally the muon
spectrometer (MS). In addition, the detector comprises a two-tier
magnet system.

In the following we will describe the basic structure of the various
ATLAS detector components (labeled 'sub-detectors'), and their spe-
cialised role in measuring different aspects of the pp collision final
state particles.

■   Magnet system

The ATLAS magnet system is designed to bend the charged outgoing
particles via the Lorentz force so as to allow for charge determina-
tion as well as a more precise momentum measurement [30]. The
innermost magnet is a thin, superconducting solenoid[17] producing a
2 TeV magnetic field [32]. This solenoid has an axial length of 5.8 m, a
radius of 2.5 m and envelops the inner detector. The solenoid magnet
was made as thin as possible in order to minimise the amount of
passive material before calorimeters.

Beyond the calorimetry, see Fig. 1.13, an additional, three-component
toroidal[18] magnet is placed. The magnet is split into a central ("bar-
rel") part, with axial length 25.3 m and radial extent between 9.4 and
20.1 m, and two end-cap toroids with a length of 5 m and outer diam-
eters of 10.7 m, serving as "plugs", optimising the bending power for
high-$|\eta|$ muons.

◼   Inner detector

The ATLAS inner detector (ID) consist of three separate parts, at increasing radial distance to the beam pipe: the pixel detector, the semiconductor tracker (SCT), and the transition radiation tracker (TRT). In combination, these three sub-detectors are designed to detect charged particles with $p_\perp$ as low as $0.5\,\mathrm{GeV}$, enable track finding throughout the entire ID envelope, and perform high-precision track momentum measurements as well as charge determination made possible by the solenoid magnet. Furthermore, each sub-detector is separated into a barrel part, with cylindrical geometry, centered at $z = 0$, and two end-cap parts, with axial symmetry, at each end of the barrel, ensuring a combined $|\eta|$-coverage up to 2.5. A cut out of the combined ID is shown in Fig. 1.14.
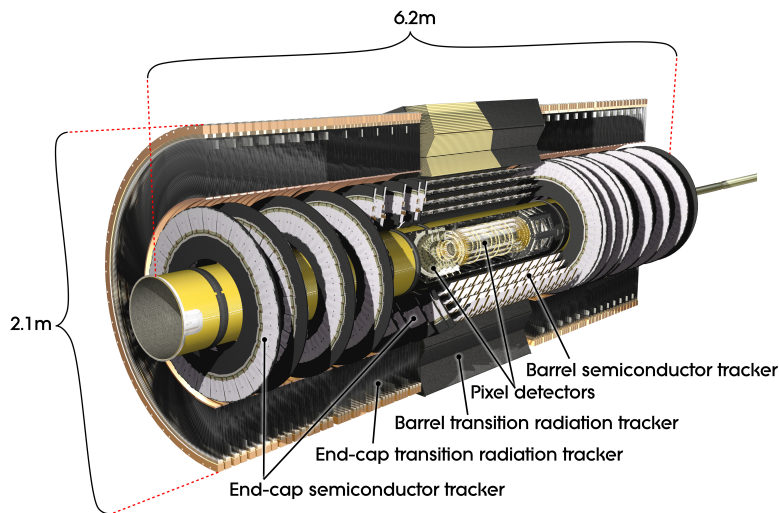
Figure 1.14: A cut out of the combined ATLAS inner detector, showing the pixel detector, SCT, and TRT as well as the rough dimensions of the ID envelope. Image from [CERN-GE-0803014-01], ATLAS Experiment © 2015 CERN.

Below we briefly describe the purpose, functioning principles, and design of each of the three sub-detectors.

Pixel detector

The pixel detector utilises semiconductor detector technology, effectively acting as solid-state ionisation chambers [34]. Operated in reverse bias, these sensors are almost completely depleted of free charge carriers. The passage of charged particles through the depletion region generates electron-hole pairs which are collected by the bias voltage, and detected as a small electric current at the read-out [30].

Such pixels enable unsurpassed spatial resolution and fast read-out, meaning that they are ideally suited for the innermost part of the detector, with a large density of particles. Here, they can measure exact coordinates for charged particles close to the interaction points, giving a robust foundation for track-finding and vertexing, see Sec. 1.4.

All sensors in the pixel detectors are identical, with a thickness of $250\,\mu$m and dimensions in $(R - \phi) \times z$ of $50 \times 400\,\mu$m^2 and corresponding intrinsic accuracies of $10 \times 115\,\mu$m^2 [32]. The choice of $1 : 8$ rectangular pixel sensors is due to the fact, that the ID is immersed in the solenoidal magnet's field, meaning that precision in the bending direction (ϕ) is more valuable than the in the axial direction—at least at small radii. The ATLAS pixel detector is segmented into a barrel part and an end-cap part separated by a 'crack' at $|\eta| \approx 2$, each with three nominal pixel layers, in order to ensure three pixel hits per track, with coverage out to $|\eta| = 2.5$. During the Long Shutdown 1 (LS1) preceding the LHC Run 2 the pixel detector had an additional pixel layer[19] (the Inner Barrel Layer [35], IBL) installed, meaning that the barrel part during Run 2 has a radial extent between $R = 31$ and $122.5\,$mm.

Semiconductor tracker (SCT)

The semiconductor tracker also uses silicon detector technology, but with a strip—rather than pixel—geometry. This allows for more affordable production of the silicon tracking sensors to cover the larger volume outside the pixel detector. However, it also gives an inherently degraded resolution, which is remedied by a pair-wise grouping of sensor modules, rotated with a small stereo angle. A charged particle crossing one SCT layer will therefore effectively traverse two silicon strips, and the hit resolution is thus given by the effective cross sectional area of the two strips.

The strips in the barrel detector have dimensions $80\,\mu$m $\times\,6.4\,$cm in $(R - \phi) \times z$, and are grouped in modules of 768 sensor strips with outer dimensions of $6.40 \times 6.36\,$cm^2 [32]. The barrel SCT has four layers, each with two sets of sensor modules of which half are oriented in parallel with the beam axis, while the other half are at a $40\,$mrad stereo angle, providing an intrinsic resolution of $17\,\mu$m $\times\,580\,\mu$m. Similar dimensions and resolutions apply for the end-cap detectors, each of which has nine layers and is separated from the barrel by a crack at $|\eta| \approx 1.3$, ensuring coverage up to $|\eta| < 2.5$.

Transition radiation tracker (TRT)

The ATLAS transition radiation tracker (TRT) relies on drift tubes for charged particle detection, in contrast to the other two ID subdetectors. These are long, gas-filled, cylindrical straw tubes (cathodes) with a tungsten wire (anode) running along their centers, operated at a large bias voltage [30]. When a charged particles passes a TRT straw, it ionises the gas atoms within the straw. The bias voltage causes the electrons to cause an avalanche of secondary ionisation with is deposited on the anode wire and read out as a signal. Furthermore, the TRT straws are interleaved with fibres or foils, which causes high-γ particles (like electrons) to emit transition radiation (TR) photons. These yield much larger signals in the TRT straws than standard minimum ionising particles, and the definition of separate low and

high thresholds (LT and HT, resp.) are used to separate electrons from e.g. pions during reconstruction (see also Sec. 1.4).

The ATLAS TRT straws have a diameter of 4 mm and an anode wire diameter of $31\,\mu$m [32], giving an intrinsic coordinate resolution of $130\,\mu$m. The barrel straws have an active lengths of 71.2 cm and are arranged axially around the beam pipe in modules which are joined at $|z| \approx 0$, comprising 73 straw layers. The end-cap detector, separated from the barrel by a crack at $|\eta| \approx 0.8$, consist of similar straws, with an active length of 37 cm, arranged in radial wheels, comprising 160 straw layers, resulting in a coverage of $|\eta| < 2.0$. The TRT straws have a much coarser resolution, compared to pixel detector and SCT, but constitute an economic way to gain tracking information over larger volumes, yielding 36 hits per track on average, and thus provides a large "lever arm" when fitting the paths of the individual tracks throughout the ID.

One disadvantage of the TRT straws, however, is the fact that they provide no coordinate information in the direction of the straws themselves. This means that the TRT barrel detector can only measure the track $R - \phi$-coordinates. However, this is a deliberate choice since, within the solenoid magnet, charged particles bend only in the ϕ direction and thus a larger number of data points are needed in order to perform a reliable fit. Since momentum of the charged particle in the z-direction is unaltered within the ID, the fewer measurements from the pixel detector and SCT suffice.

Energy resolution

In the homogeneous, solenoidal field inside the ID, charged particles will (ignoring scattering and other energy losses) move on a helical trajectory. Each signal in the various ID sub-detectors correspond to measurements of this path, which can be combined to form a "track" left by the charged particles (see Sec. 1.4). The combined ID measures the momentum of charged particles based on the curvature of this helix. The main uncertainty on the momentum measurement based on tracking is given by the hit position measurements [30]. These give a relative $p_\perp$ resolution of

$$\frac{\sigma_{p_\perp}}{p_\perp} \propto p_\perp \qquad (1.44)$$

■ Calorimetry

Charged particles with $p_\perp > 450$ GeV have sufficient energy to pass through the ID to the calorimetry that envelopes it. Similarly, neutral particles pass through the ID, completely undetected, however, and with no deflection caused by the solenoidal field. The ATLAS calorimetric system is designed to measure their direction and energy. This system can, like the ID, be grouped into three parts[20]: the electro-magnetic (ECAL), hadronic (HCAL), and forward calorimeter (FCAL). The combined system is designed to proved precision electromagnetic (EM) measurements with an $|\eta|$ coverage similar to that of the ID,

[20] Although this assignment is not at all unambiguous.

as well as measurements of the $\eta - \phi$ coordinates and total energy carried by neutral final-state particles, not measured by the ID, over a large $|\eta|$ range so as to minimise longitudinal leakage. The ECAL and HCAL have geometries like the ID sub-detectors, each with a barrel and two end-cap components. The FCAL is a purely forward sub-detector and thus has no barrel part. A cut out of the combined calorimeter system is shown in Fig. 1.15.

Figure 1.15: A cut out of the ATLAS electromagnetic (ECAL), hadronic (HCAL), and forward (FCAL) calorimeters. Image from [CERN-GE-0803015], ATLAS Experiment © 2015 CERN.
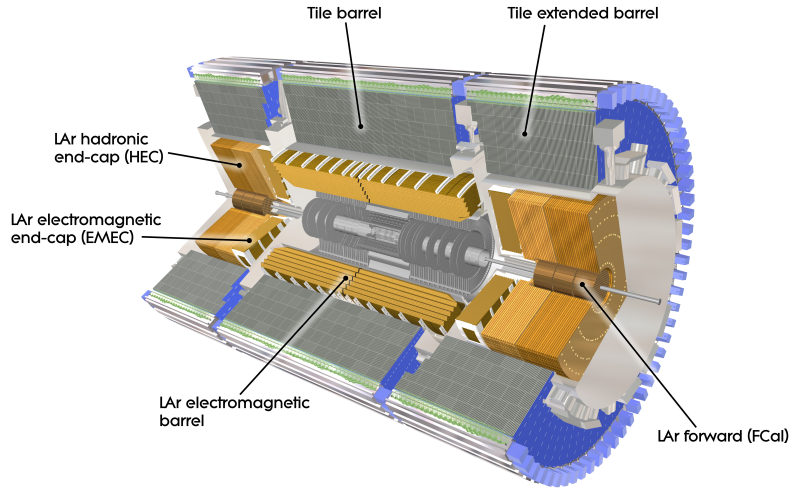
All components of the ATLAS calorimeter system are sampling calorimeters, where a passive material, generating cascades or showers of particles, is interleaved with active material, measuring the number of particles in the cascade. The ECAL and HCAL are furthermore segmented in depth[21] in order to provide measurements of the shower shape (used in particle identification, see Sec. 1.4) and better pointing information in $\eta - \phi$.

[21] Radially for barrel components, and axially for end-caps.

Electromagnetic calorimeter (ECAL)

Electron and photons interact differently from hadrons, having no strong interaction, and can thus be detected with specialised calorimetry. Electrons lose energy to EM bremsstrahlung, loosing a factor of $1/e$ of their energy over one radiation length (X_0) [30]. These bremsstrahlung photons convert to e^+e^- pairs after $9X_0/7$, thus giving rise to an EM shower. The same is true for final-state photons which, however, start their shower slightly later and correspondingly have slightly different maximal shower depths.

The ATLAS ECAL uses lead ($X_0 = 5.6\,\mathrm{mm}$) as its passive material and liquid argon (LAr) as the active medium, in which electrodes are embedded for signal detection. It uses an accordion geometry, see Fig. 1.16, designed to ensure full ϕ coverage and this no leaks from particles escaping the potential gaps between detector modules.

The ECAL envelops the ID, with a coverage of $|\eta| < 1.475$ ($1.375 < |\eta| < 3.2$) in the barrel (end-cap) component [32], where the end-cap
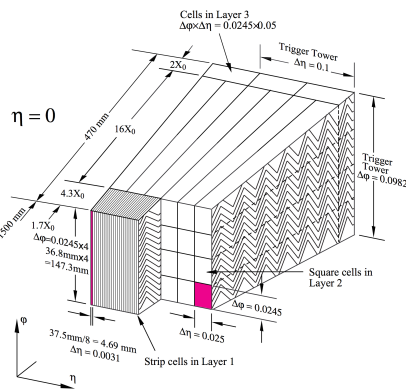

Figure 1.16: Diagram of a section of the ATLAS barrel ECAL, showing the according geometry, the three depth segments, as well as the different $\eta - \phi$ granularity across the layers. Image from [32].

is segmented into two concentric, axially symmetric wheels (outer and inner). The $|\eta| < 2.5$ range, matching the ID, is dedicated to precision physics as has three depth layers[22], while the remaining ECAL (the inner end-cap wheel) has two. These three layers are segmented in $\eta - \phi$, the first layer ("strips") with a fine $\Delta\eta = 0.003$ resolution, for precise η pointing, the remaining layers with gradually coarser, square segments, providing a additional lever arm in the $\eta - \phi$ measurement. The granularity of the ECAL varies across layers and $|\eta|$, but is generally on the order of $\Delta\eta \times \Delta\phi \gtrsim 0.025 \times 0.025$. The barrel ECAL has a total depth of $\sim 22\,X_0$ [32], where the second layer collects the majority of the deposited energy, and the third layer is designed to collect the tails of the EM showers and determine any punch-through into the HCAL.

Hadronic calorimeter (HCAL)

Final-state hadrons leave only a fraction of their energy in the ECAL, so the HCAL relies on strong interaction to measure their energy and stop them from escaping the detector. The length scale of process is characterised by the nuclear interaction length (λ), which is the mean free path between inelastic strong interactions [30]. Each interaction produce secondary particles (mainly pions) which themselves interact with the passive medium, giving rise to a hadronic shower.

The HCAL is divided into a barrel and extended barrel component, covering $|\eta| < 1.0$ and $0.8 < |\eta| < 1.7$, resp., as well as two end-caps covering $1.5 < |\eta| < 3.2$ [32]. The (extended) barrel components use lead ($\lambda = 17.6\,\text{cm}$) interleaved with scintillating tiles as active material. These components are constructed from 64 wedge modules, each with three depth segments, placed in an axially symmetric fashion around the ECAL for an almost hermetic construction, each wedge spanning $\Delta\phi \approx 0.1$. The (extended) barrel components have granularities of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ in the first two layers and $\Delta\eta \times \Delta\phi = 0.2 \times 0.1$ in the last.

The end-caps are built from 32 wedges, each made up of copper plates (passive) interleaved with LAr for sampling. Each end-cap consists of two independent wheels, placed in succession along the beam axis, with two layers each for a total of four depth layers (along z) with granularities of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ (0.2×0.2) in the $1.5 < |\eta| < 2.5$ ($2.5 < |\eta| < 3.2$) region.

In the barrel region, the entire calorimeter system has a depth of $9.7\,\lambda$, of which the HCAL makes up $7.4\,\lambda$. The large total depth of the combined calorimeter system means that there is minimal punch-through to the muon spectrometer and that the calorimeter system provides reliable measurements of the *total* energy in the event, needed for reconstructing any missing energy, see Sec. 1.4.

Forward calorimeter (FCAL)

In order to ensure high-$|\eta|$ coverage (needed for reconstructing missing transverse energy, see Sec. 1.4), the ATLAS calorimeters system

[22] An additional presampler (PS) is positioned before the barrel ECAL within $|\eta| < 1.8$, designed to measure the energy lost before the ECAL.

also includes a forward calorimeter ($3.1 < |\eta| < 4.9$), ensuring a combined calorimetric coverage of $|\eta| < 4.9$ [32]. The FCAL has three modules (along z), each with a depth of $45\,\mathrm{cm}$: the first, intended for mostly EM measurements, uses copper as absorber; the second and third layer intended for hadronic measurements, use tungsten (high λ). All three layers are constructed as matrices of tubes, parallel to the beam axis, each tube containing a metal rod and LAr in the gaps between them as active material. The HCAL has a total depth of $\approx 10\,\lambda$, thus offering excellent containment even at high $|\eta|$.

Energy resolution

Since all components of the ATLAS calorimeter system are sampling calorimeters, the energy measurements in each effectively amount to a counting (by the active material) of the number of particles produced in the cascades [30]. Therefore, the main energy resolution of the calorimeters are given by a Poissonian term, meaning that the relative resolution goes as

$$\frac{\sigma_E}{E} \propto \frac{1}{\sqrt{E}} \tag{1.45}$$

This means that for charged particles with large momenta ($p_\perp \gtrsim 25\,\mathrm{GeV}$ for electrons, [30]), the resolution in the calorimeters becomes better than that from tracking in the ID.

■ Muon spectrometer

The only particles escaping the calorimeter system[23] are muons, due to their not participating in strong interactions and having a large mass[24].

The ATLAS muons spectrometer (MS) relies on the bending of muons within the large volume of the air-core toroid magnet to achieve a large lever arm for excellent stand-alone momentum determination and charge identification for muons with as high as $p_\perp \gtrsim 3\,\mathrm{TeV}$ [32]. In order to perform these measurements, the MS is composed of four separate sub-detectors, with a general barrel-and-end-cap structure like the ID and calorimeters. A cut out of the combined muon spectrometer is shown in Fig. 1.17.

The monitored drift tube (MDT) chambers function similarly to the TRT straws, but have a diameter of $30\,\mathrm{mm}$ and are positioned in modules perpendicular to the $R - z$ plane, i.e. the bending plane of the toroidal magnet field. This means than, similarly to the TRT, the MDT can only measure the track coordinate in the bending direction (here, η). In the barrel region ($|\eta| < 1.05$) the MDT has three layers at radii $R \approx 5$, 7.5, and $10\,\mathrm{m}$. The end-cap region ($1.05 < |\eta| < 2.7$) is constructed from four axially symmetric wheels, positioned at $|z| \approx 7.4$, 10.8, 14, and $21.5\,\mathrm{m}$.

In the inner-most layer of the end-cap's $2.0 < |\eta| < 2.7$ region, the MDT is replaced by the cathode strip chamber (CSC) sub-detector, due to considerations regarding the rate[25]. The CSC consists of pairwise orthogonal multi-wire proportional chambers in four consecutive

[23] Apart from neutrinos, which only interact weakly and thus are never detected, meaning that they only show up as missing energy otherwise expected from momentum conservation in the transverse plane. See sec. 1.4 for details.

[24] Due to its high mass, $m_\mu/m_e \approx 200$ [2], since bremsstrahlung is suppressed by $(m_e/m_\mu)^2$ [36].

[25] The MDT has drift times of up to $700\,\mathrm{ns}$ [32].

layers enabling the sub-detector to measure both η and ϕ coordinates for each track.

Finally, the MS includes the resistive-plate (RPC; parallel electrode-plate detectors, with metal strip read-out) and thin-gap chambers (TGC; multiwire proportional chambers) positioned in the barrel and end-cap regions, resp. These are mainly intended for triggering, but also provide measurements of the second track coordinate (ϕ), which is assigned to the MDT measurements of the corresponding track.

Tab. 1.3 shows the nominal resolution and average expected number of measurements per track for each of the four MS sub-detectors.

| Sub-detector | Resolution in | | Measurements / track | |
	z/R	ϕ	barrel	end-cap
MDT	$35\,\mu$m (z)	–	20	20
CSC	$40\,\mu$m (R)	$5\,$mm	–	4
RPC	$10\,$mm (z)	$10\,$mm	6	–
TGC	2–$6\,$mm (R)	3–$7\,$mm	–	9

Table 1.3: Resolution and average number of measurements per track for each sub-detector in the ATLAS muon spectrometer. Adapted from [32].

Muons are also detected in the ID and therefore the measurements from the MS are usually combined with those from the ID in order to benefit from the high-resolution measurement points close to the IP.

Energy resolution

The large lever arm in measuring the muon momentum in the MS has the associated disadvantage of introducing the possibility of multiple scattering in the larger volume. This is the reason for using an air-core toroid magnet outside the calorimeter system, thus minimising the probability for muons to scatter. The momentum resolution due to

multiple scattering is given by [30]

$$\frac{\sigma_{p_\perp}}{p_\perp} \propto \text{const.} \tag{1.46}$$

This means that at low $p_\perp$ this constant uncertainty dominates the muon momentum resolution, while at high $p_\perp$ a tracking-based uncertainty similar to that in Eq. (1.44) takes over.

■ Design performance

The combined ATLAS detector as designed to meet stringent performance requirements, necessary to fulfill its physics goals, see [32]. These requirements can be summarised as in Tab. 1.4, describing the combined coverage in $|\eta|$ of each detector system as well as its nominal $p_\perp$ resolution.

Table 1.4: Performance goals of the ATLAS detector. "⊕" means sum in quadrature. All values of $p_\perp$ and E are in units of GeV. For high-$p_\perp$ muons the MS performance is independent of the ID. Adapted from [32].

Component	Required resolution			η coverage	
ID	$\sigma_{p_\perp}/p_\perp =$	$0.05\% \, p_\perp$	$\oplus \ 1\%$	$0 < |\eta| < 2.5$	
ECAL	$\sigma_E/E \ =$	$10\% \ / \sqrt{E}$	$\oplus \ 0.7\%$	$0 < |\eta| < 3.2$	
HCAL					
barrel/e.c.	$\sigma_E/E \ =$	$50\% \ / \sqrt{E}$	$\oplus \ 3\%$	$0 < |\eta| < 3.2$	
forward	$\sigma_E/E \ =$	$100\% \ / \sqrt{E}$	$\oplus \ 10\%$	$3.1 < |\eta| < 4.9$	
MS	$\sigma_{p_\perp}/p_\perp =$	10%	at $p_\perp = 1\,\text{TeV}$	$0 < |\eta| < 2.7$	

The energy resolutions are seen to follow the general behaviours anticipated above. We will use these nominal performances later, in Sec. 3.1, when we perform a crude simulation of the ATLAS detector.

1.4   Reconstruction

We have now described purpose, design, and nominal performance of the different sub-detectors of the ATLAS experiment. From their measurements, we have to reconstruct, to the best of our abilities, the actual particles emanating from the interaction point. Therefore, we begin by discussingtrack reconstruction and calorimeter clustering, which will serve as the basis for the identification of different species of particles. Finally, we will turn to the more involved matters of reconstructing vector bosons and the concept of "jets".

■   Inner detector

As described in Sec. 1.3, the passage of charged particles through the ID (and MS) result in a number of "hits" with varying degrees of precision: from "space-points" [37] in the pixel detector ($\sigma \gtrsim 10\,\mu$m) to measurements of single coordinates in the TRT and MDT ($\sigma \approx$ 130 μm and 35 μm, resp.). These hits will[26] fall on a helix path in the solenoidal magnetic field inside the inner detector, and then bend in the $R - z$ direction in the toroidal field outside the ID envelope.

[26] Modulo energy loss and multiple scattering.

  Our current task is now to combine these hits into a "track" traversing the detector as well as using the precision measurements of the pixel detector (in particular the IBL) to determine the interaction point from which the track originated, as there may be several cf. the mentioning of pile-up in Sec. 1.2.

Track reconstruction

The silicon detectors (pixel and SCT) perform the most accurate measurements of charged particle coordinate, and each hit is represented as a three-dimensional "space-point" [37], either from the barycenter [30] of a pixel cluster (since a single charged particle may give hits in multiple adjacent pixels [38]) or from that of two overlapping SCT stereo strips. Nominally, three or more space points aligning with the interaction point are used as seeds for the tracking algorithm [37], which iteratively adds hits outwards from the seeds, and into the TRT, using local pattern recognition based the Kalman filter technique [39]. The resulting trajectory is not necessarily completely smooth, due to the step-wise nature of the Kalman filter, which allows for inclusion of kinks due to multiple scattering. If the track reconstructed in this manner satisfies some quality criteria (e.g. regd. $p_\perp$[27], impact parameters wrt. the beam spot[28], or the number of silicon hits), the track is stored, and the remaining (unassigned) hits are used for reconstruction additional tracks.

[27] Only tracks with $p_\perp \gtrsim 500\,$GeV are reconstructed [32].

[28] The center of the interaction region, known in advance from vertex positions in previous bunch crossings [40].

Momentum measurement

The momentum of a charged particle is determined by measuring the bending of the corresponding track in the inner detector, specifically using the sagitta method [30]. In the bending plane ($R - \phi$) the track

trajectory will be a circular segment with radius

$$\rho = \frac{p_\perp}{qB} \tag{1.47}$$

where $p_\perp$ is the track momentum in the bending plane, q in the charge of the particle, and B is the (homogeneous) field strength. This relation can similarly be expressed in terms of the sagitta[29] s, which is operationally easier to measure, giving

$$p_\perp \approx \frac{qL^2B}{8s} \tag{1.48}$$

where L is the shortest distance between the end points of the trajectory defining the sagitta. Note here, that the particle charge q is not known in advance and must therefore be identified/assigned. The actual measurement is therefore actually of $p_\perp/q$, or its inverse.

Vertex reconstruction

The collection of tracks reconstructed in this way now form the basis for reconstructing production vertices along the beam line. All track compatible with origination from the interaction region (given their impact parameter errors) are included in the determination of the global maximum of z coordinates, used as seed for the vertex fitting algorithm [40]. The algorithm includes all tracks within 7σ of the iteratively fitted vertex z coordinate, requiring at least two associated tracks for reconstructing a vertex. All tracks not compatible with the vertex are used to seed another vertex finding, and so on until no more vertices are found. In this way it is possible to reconstruct several vertices, corresponding to the hard interaction (the one with the highest sum $p_\perp$ of all associated tracks) as well as a number of pile-up interactions. Finally, a number of tracks may not be associated to any vertex. The implications of this is discussed further in Sec. 4.1.

■ Calorimeters

Outside the ID, particles deposit their energy in the cells of the calorimeter system. Since all particles (except the muon) initiate a shower with a considerable lateral extent (e.g. the Moliere radius in the lead of the barrel ECAL is $R_M \approx 1.3\,\mathrm{cm}$ [36], excluding upstream bremsstrahlung photons, compared to a layer 2 cell size of $(R - \phi) \times z \approx 3.7 \times 3.7\,\mathrm{cm}^2$ [32]) we need a prescription for association multiple cells into clusters believed to represent individual particles.

EM clusters

Towers in the ECAL are build from the cells in all longitudinal layers (PS, strips, middel, back) within windows[30] of $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$ by summing the energy depositions in each layer [41]. "Precluster" seeds are the found by moving a window covering a fixed number of $N_\eta \times N_\phi = 5 \times 5$ towers across the ECAL and computing the barycenter of all regions where the sum energy within the window

is a local maximum and above some threshold ($\sim 3\,\text{GeV}$). From these seeds, EM clusters are formed in succession in each ECAL layer, using fixed-size windows of size[31] $N_\eta \times N_\phi \sim 3 \times 5$, around the barycenter of the energy depositions within the window in (one of) the previous layers: the precluster barycenter seeds the window in the middel layer; the barycenter in the middle layer, seeds the window in the strip and back layer; while the barycenter in the strip layer seed the window in the PS.

TopoClusters

Hadrons do not deposit all of their energy in the ECAL, and correspondingly we need to include the HCAL and FCAL in clustering their energy. This is done by topological clustering [41], which results in variably sized clusters in contrast to the fixed-size EM clusters.

The clustering procedure searches all cells in the combined calorimeter system, and store all those with an energy $|E| > 4 \times \sigma$ as "protoclusters", where σ is the total expected noise[32] in a particular cell. All neighbouring cells (both within and across sub-detectors and layers) with energies $|E| > 2 \times \sigma$ are added to the adjacent proto-cluster until no such cells remain. If, in the clustering process, two proto-clusters share a neighbouring cell with $|E| > 2 \times \sigma$, they are merged into one. Finally, all cells ($|E| > 0 \times \sigma$) at the edge, in all directions, of a proto-cluster are added to it. This clustering is called the "420" scheme, due to the energy thresholds used.

In order to account for the possibility of energy depositions from distinct particles overlapping in the calorimeters, after the merging, local maxima cells[33] are found within each proto-cluster, requiring the cell to have $E > 500\,\text{MeV}$, to have ≥ 4 neighbouring cells, and to have an energy larger than any those. The clustering is then repeated—only within the parent proto-cluster—without any requirements in the neighbouring cell energies. If, in the process, two forming protoclusters share a cell, the two proto-clusters are not merged, but instead the contribution from the cell is split between them, with a weight given by their relative energies and the distance from the cell to the center of each proto-cluster.

The proto-clusters left after this procedure are then promoted to topological clusters, or TopoClusters; three-dimensional energy deposits in the combined calorimeter system, which may share cells on their boundaries.

Particle flow

The calorimeter clusters, found as described above, have three problems compared to tracks:

1. charged particle are deflected in the ID solenoidal magnetic field, meaning that their calorimeter clusters do not reflect their "true" ϕ value (i.e. the one at the particle level at the interaction point);

2. even for netural and high-$p_\perp$ charged particles (whose deflection

[31] Rectangular windows are used due to the deflection and ensuing bremsstrahlung of charged particles and the subsequent larger spread in solenoid bending direction (ϕ).

[32] This is computed as the sum in quadrature of the total electronics noise and the expected contribution from pile-up.

[33] By default, cells in the HCAL and the PS and strip layer of the ECAL are excluded, in order to suppress noise. However, if no maxima are found in the middle and back layers of the ECAL which overlap in η and ϕ with maxima found in the excluded layers, these maxima may be included as well, in order to account for hadronic clusters with minimal EM activity.

is negligible, cf. Eq. (1.47)), clusters do not have high-precision measurements of their pointing, thus preventing the measurement of the particles' z-coordinate and, therefore, the determination of their origin vertex;

3. for charged particles with $p_\perp \lesssim 100\,\mathrm{GeV}$, their energy measurement is poorer (in the HCAL) than that in the ID, cf. Tab. 1.4.

To this end, we may use the ID track information to improve the calorimetric measurements, which ATLAS has done in Run 2 with the implementation of 'particle flow' ("PFlow") through the `eFlowRec` algorithm [42]—something which has been used by CMS since Run 1 [43].

The `eFlowRec` algorithm takes as input the lists of reconstructed tracks[34] and TopoClusters. It this extrapolates all track trajectories to their expected impact with the second layer of the ECAL. Using the extrapolated track coordinates $(\eta_{\mathrm{trk.}}, \phi_{\mathrm{trk.}})$, it then searches for the nearest TopoCluster (TC) in terms of

$$\sqrt{\frac{(\eta_{\mathrm{trk.}} - \eta_{\mathrm{TC}})^2}{\sigma_\eta^2} + \frac{(\phi_{\mathrm{trk.}} - \phi_{\mathrm{TC}})^2}{\sigma_\phi^2}} \tag{1.49}$$

similar to the measure in Eq. (1.35), but weighted by the standard deviation of the TopoCluster coordinates $\sigma_{\eta,\phi}$. If the nearest TopoCluster satisfies[35]

$$E_{\mathrm{TC}} > E_{\mathrm{trk.}} - k \times \sigma_{\mathrm{trk.}} \tag{1.50}$$

that is, that the cluster energy is compatible with the track energy, within some tolerance, then the track energy is subtracted from the TopoCluster, and the track is stored as a (charged) PFlow object. If the modified cluster has an energy which is compatible with zero within the expected fluctuations from the subtraction ($\propto \sigma_{\mathrm{trk.}}$) it is discarded; otherwise it is kept as a (neutral) PFlow object. This process is carried out for all tracks, after which `eFlowRec` performs 'Split Shower Recovery' on the tracks failing the requirement in Eq. (1.50): it compares the track energy to the sum energies for all clusters with barycenters within a cone of radius $dR < 0.2$, in order to include particles whose calorimeter showers are split into multiple TopoClusters, such that either of these individually fails the requirement in Eq. (1.50).

The output of the algorithm is a list of PFlow objects, comprehensively incorporating information from the ID and calorimeter system, which contains:

charged PFlow objects, covering

- isolated tracks,
- tracks matched to TopoClusters (through subtraction).

neutral PFlow objects, covering

- isolated TopoClusters,
- TopoClusters with energy modified by track energy substraction.

Note that, as such, particle flow is only performed for TopoClusters, and not for EM clusters.

[34] Only tracks with $0.5\,\mathrm{GeV} < p_\perp < 40\,\mathrm{GeV}$ are used, since softer tracks are not reconstructed, while harder tracks will experience negligible deflection and have superior energy measurement in the calorimeter system.

[35] In the present presentation we ignore complications regarding calibration.

■ Particle identification

We have new defined sets of clusters, tracks, or PFlow objects, each with definite coordinates. Our task is then to identify these objects with physical particles. Since our main concern is the $WW \rightarrow \ell\nu q\bar{q}$ final state, cf. Sec. 1.1, with[36] $\ell = e, \mu$, we will start by focusing on the identification of electrons, muons, and neutrinos.

Electrons

The detector signature left by electrons is a track in the ID followed by a shower in the ECAL which a large degree of containment [37]. Therefore, electron candidates are initially reconstructed [44] in the barrel region ($|\eta| < 2.5$) as an EM cluster matched to a ID track[38] (extrapolated to the second layer of the ECAL) within[39] $|\Delta\eta| < 0.05$ and $|\Delta\phi| < 0.1$. The energy of the electron candidate is calculated from four components[40]: the energy measured in the cells of the EM cluster, and the estimated energies deposited before the ECAL, after the ECAL (longitudinal leakage), and outside the cluster (lateral leakage). The electron candidate's η and ϕ coordinates are taken from the matched track at its production vertex.

   Having found a set of electron candidates, we impose a selection designed to retain true, isolated electrons while rejecting backgrounds from e.g. hadron and photon misidentification. This selection is based on a series of cuts regarding the shape of the EM shower, the leakage into the HCAL, and additional requirements on the quality of the track. These cuts come in different degrees of "tightness", labeled *loose*, *medium*, and *tight*, to be used in analyses, reflecting the ability of the selection to reject background.

   Beyond $|\eta| = 2.5$ there is no tracking aid, and electron candidates are found only from clusters in the ECAL end-cap and the first layer of the FCAL. Correspondingly the identification also differs from that for central electron. We will, however, only be interested in central electrons.

Muons

Muons have a very clean detector signature, being the only (measurable) particle to escape the calorimeter system. We thus expect muons to show up as tracks in both the ID and the MS. The reconstruction of 'combined' muons [45] therefore computes the χ^2 deviation[41] between pairs of tracks in the ID and MS. The combined track is then found by a statistical combination of the matched ID and MS tracks, extrapolated to the point of closest approach to the beam line, weighted by the sum of covariance matrices for each track. A set of quality criteria exist for the muons, similarly to those for the electron above.

Neutrinos

Since neutrinos only interact through weak interaction, they will escape the experiment undetected. However, the initial state partons

[36] Since τ-leptons are notoriously tricky to reconstruct, we leave them out of the current search.

[37] Very little punch-through into the HCAL, due to the $\sim 22\,X_0$ depth of the ECAL, cf. Sec. 1.3.

[38] EM clusters without associated tracks are considered (unconverted) photon candidates.

[39] The asymmetry is introduced in order to account for the emission of bremsstrahlung photons in the bending direction.

[40] For electrons with $p_\perp \gtrsim 30\,\mathrm{GeV}$, the energy measurement in the ECAL is more precise than that in the ID, cf. Tab. 1.4.

[41] From the covariance matrix for the five perigee parameters [30]: signed distance of closest approach to beam line, d_0; z-coordinate at point of closest approach; azimuthal angle of the track tangent, ϕ_0; polar angle of the track tangent wrt. the z-axis, θ_0; and charge over momentum, q/p, cf. Eq. (1.48).

of a hard interaction resulting in final state neutrinos are parallel with the z-axis (modulo ISR). This means that they have no net momentum in the transverse plane and, by momentum conservation, we therefore expect the vector transverse momenta of final state particles to also sum to zero. This clear view of things is naturally complicated by pile-up collisions, calorimeter energy calibration issues, etc., but since the momentum of the neutrino is not measured, its detector signature will be the presence of a significant missing transverse energy, defined as

$$\mathbf{E}_\perp^{\text{miss}} \equiv - \sum_{i \in \text{TopoClusters}} (p_x{}^i, p_y{}^i), \qquad E_\perp^{\text{miss}} = |\mathbf{E}_\perp^{\text{miss}}| \qquad (1.51)$$

Since particles will always escape the calorimeter coverage along the beam line, a similar relation does not exist for the longitudinal momentum of the neutrino.

■ Reconstructing vector bosons

Having described the reconstruction and identification of the particles in our final state of choice, we now turn to the reconstruction of the W bosons themselves: one decaying leptonically and one hadronically.

Leptonic decay

The boson decaying through $W \to \ell \nu$ would ideally be reconstructed as the four-vector sum of the lepton and the neutrino. However, since the z-component of the neutrino momentum cannot be determined from measurement, we decide to estimate it—provided we have identified a high-$p_\perp$ lepton and significant $E_\perp^{\text{miss}}$—under the constraint that the invariant mass of the combined $\ell \nu$ system should be equal (or as close as possible) to the W boson pole mass, $m_W = 80.385\,\text{GeV}$ [2]. The computation[42] is done in App. A.1 and the result is

$$\eta_\nu = \eta_\ell \pm \text{arccosh} \left(\frac{m_W^2}{2 p_{\perp \ell} p_{\perp \nu}} + \cos \Delta\phi \right) \qquad (1.52)$$

where $\Delta\phi$ is the azimuthal angle between ℓ and $\mathbf{E}_\perp^{\text{miss}}$. If the argument of arccosh is ≥ 1 this function is well-defined and we choose the sign of the arccosh-term that makes $|\eta_\nu|$ minimal. This is justified since the two signs are equally valid, but due to the pseudo-rapidity plateau structure of events, cf. Sec. 1.2, the smallest value of $|\eta_\nu|$ is the most probable. If the argument is < 1, we choose $\eta_\nu = \eta_\ell$. See App. A.1 for details.

Since we have $p_{\perp \nu} \equiv E_\perp^{\text{miss}}$ and $\phi_\nu \equiv \arg(\mathbf{E}_\perp^{\text{miss}})$ from measurement, and identically set $M_\nu = 0$, this completely specifies our estimate for the neutrino four-vector. The leptonically decaying W boson is then, as noted, given by the four-vector sum $W^\mu = \ell^\mu + \nu^\mu$.

Hadronic decay

The boson decaying through $W \to q\bar{q}$ will not be as clearly defined as the leptonic one, since the two (highly energetic) final state quarks will

[42] Just for fun it is done in a slightly different way that the standard quadratic equation in $p_{\nu,z}$, but the results should be equivalent.

initiate parton showers which will subsequently decay to a collection of hadrons, cf. Sec. 1.2. This collection is harder to identify as a W boson than a $\ell\nu$ system, since hadronically decaying Z bosons and—to an overwhelming degree—non-resonant, high-$p_\perp$ partons from purely QCD processes result in seemingly identical final states.

The spray of particles resulting from the splitting of a hard parton will be collimated around the direction of the original parton due to the collinear divergence in Eq. (1.39) as well as the conservation of momentum in the plane transverse to the direction of the original parton. Therefore, the spray of particles will be confined to relatively small area of the detector, which we can reconstruct as a "jet" of particles.

A W boson decaying at rest will therefore result in two (low-$p_\perp$) jets in opposite hemispheres of the detector; one for each parton. However, bosons originating from a high-mass resonance decay will themselves be boosted, with $p_\perp \approx M_R/2$, meaning that the final state quarks, and their reconstructed jets, also will be boosted in the direction of the W, with a separation of [46]

$$dR_{qq} \approx dR_{jj} \approx \frac{M_W}{p_\perp} \frac{1}{\sqrt{z(1-z)}} \xrightarrow[z \approx 0.5]{} \frac{2M_W}{p_\perp}; \quad p_\perp \gg M_W \quad (1.53)$$

where z is the momentum fraction of one of the quarks and $p_\perp$ is the boson transverse momentum. This means, that if the W boson is sufficiently boosted, the distance between the two individual quark jets, dR_{jj}, may be so small that they overlap, such that they will be reconstructed as a single, fat jet J. The main features distinguishing boosted boson jets from non-resonant QCD jets will therefore be their masses (expected to resemble the pole mass of the boson in question) and the fact that they may have some residual "two-prong" structure, even when reconstructed as a single jet, which we may exploit.

■ **Jets**

In order to identify a (boosted) boson jet, we therefore need some jet definition. This is done using one of a selection of jet clustering algorithms.

Clustering algorithms

The jet clustering algorithms used at present-day experiments are all based on sequential recombination [47]. Since the colour-correlations in an event are not just between the hard, final state partons but also between these and the beam remnants, cf. Sec. 1.2, the sequential recombination algorithms define two distance measures:

$$d_{ij} = \min(p_{\perp i}^{2p}, p_{\perp j}^{2p}) \frac{dR_{ij}^2}{R^2} \quad (1.54a)$$

$$d_{iB} = p_{\perp i}^{2p} \quad (1.54b)$$

where dR_{ij} is the rapidity-based distance measure between to particles i and j, cf. Eq. (1.33), and $p_{\perp\{i,j\}}$ is their transverse momenta. Here,

Eq. (1.54a) is a measure of the distance between particles i, j and Eq. (1.54b) is a measure of the closeness of particle i to the beam. R is called the distance parameter of the algorithm and p is a an algorithm-specific parameter: the $k_\perp$ algorithm [48] uses $p = 1$, the Cambridge-Aachen algorithm [49] uses $p = 0$, and the anti-$k_\perp$ algorithm [50] uses $p = -1$. Starting from a collection of reconstructed particles (for ATLAS, the set of all TopoClusters), each of the algorithms then proceeds as shown in Fig. 1.18 [47].

Figure 1.18: Pseudo-code implementation of the sequential recombination jet clustering algorithms.

```
1   given particleCollection , distanceMeasures
2   jetCollection = { }
3   while particleCollection .size > 0 do
4       compute all dᵢⱼ and dᵢB using distanceMeasures
5       if min({dᵢⱼ}) < min({dᵢB}) then
6           remove i and j from particleCollection
7           add (i + j) to particleCollection
8       else
9           remove i from particleCollection
10          add i to jetCollection
11      end
12  end
```

These algorithms sequentially recombine particles, which are close according to some distance measure Eqs. (1.54), by four-momentum summation until no more nearby particles are found. The distance parameter R determines when the recombination stops, and can therefore be seen as a proxy for the radius of the resulting jet[43].

By their different choices of p, we see that the $k_\perp$ algorithm favours clustering soft (and close) particles first, while the anti-$k_\perp$ favours clustering hard (and close) particles first and then adding the softer contributions later, and the Cambridge-Aachen algorithm is seen to be purely geometric.

[43] This is mostly true for the anti-$k_\perp$ algorithm [46], which yields jets which are approximately circular.

Jet grooming algorithms

The concept of jets is a construct and no algorithm will perfectly recombine all the final-state hadrons from a particular hard interaction parton, and only these. As a result, soft physics contamination from the underlying event and pile-up are bound to be included in any jet reconstructions. This has the consequence of degrading the energy and mass measurements, which we rely on for distinguishing resonant boson jets from non-resonant QCD jets, see above. Therefore, a range of techniques have been proposed to "groom" jets, i.e. remove soft components while keeping the hard constituents (and therefore, it is hoped, the contribution from the hard interaction itself) intact. Below, we will quickly describe the three predominant grooming schemes used at LHC experiments:

Mass-drop filtering [51], or 'BDRS' after the authors, searches for the two-prong structure associated with boosted boson decay. It takes

as input a sequentially clustered jet, which is then iteratively decomposed by undoing the previous step of the recombination: $j \rightarrow j_1$ and j_2, such that $M_{j_1} > M_{j_2}$. Two conditions are imposed on the mass drop and the $p_\perp$ asymmetry:

$$\frac{M_{j_1}}{M_j} < \mu \qquad (1.55\text{a})$$

$$y = \frac{\min(p_{\perp j_1}^2, p_{\perp j_2}^2)}{M_j^2} dR_{j_1 j_2}^2 > y_{\text{cut}} \qquad (1.55\text{b})$$

where μ and y_{cut} and free parameters (typically 0.67 and 0.09, but 1.00 and 0.04 for the modified BDRS-A method, which we will use in Secs. 3.1 and onwards). If both conditions are met, the jets $j_{1,2}$ are deemed to represent a symmetric two-prong system and j is kept as the output jet. Otherwise, if either condition is not met, j_2 is deemed too soft and/or collinear to represent a hard prong and is discarded, j_1 is labeled j, and the decomposition is continued. The constituents of the output jet j from the mass-drop procedure is then reclustered using a smaller distance parameter R_{filt} (usually 0.3), keeping only the three hardest jets (henceforth called "subjets"), the hardest two describing the $W \rightarrow q\bar{q}$ quarks, the third allowing for the possibility for one hard gluon emission. The two steps of the mass-drop filtering procedure are show in Fig. 1.19.

j_2

j_1

Initial jet $\qquad m^{j_1}/M^{\text{jet}} < \mu_{\text{frac}}$ *and* $y > y_{\text{cut}}$
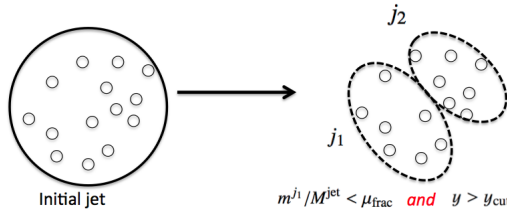
(a) The mass-drop step of the BDRS method.

Initial jet $\qquad R_{\text{filt}} = \min[0.3, \frac{\Delta R_{j_1, j_2}}{2}] \qquad$ Filtered jet
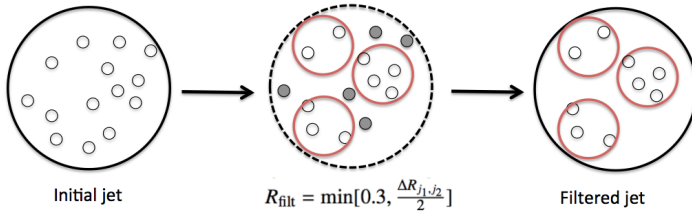
(b) The filtering step of the BDRS method.

Figure 1.19: Diagrammatic representation of the mass-drop filtering method for jet grooming. Adapted from [52].

Trimming [53], also uses sequentially clustered jets as input, but instead of retracing the clustering sequence, it reclusters the constituents of the jet with a smaller radius parameter $R_{\text{sub}} < R_{\text{jet}}$ (usually 0.3). Each such clustered jet is then compared to the original jet, and only the ones carrying a significant fraction (usually 0.5) of the original jet $p_\perp$ are kept, i.e. the ones with:

$$p_{\perp i} > f_{\text{cut}} \, p_\perp \qquad (1.56)$$

are discarded, where $p_{\perp i}$ is the transverse momentum of the i'th "subjet", and $p_\perp$ is that of the original jet. The remaining jets are then combined into a trimmed jet. The procedure is shown in Fig. 1.20.

Figure 1.20: Diagrammatic representation of the trimming method for jet grooming. Adapted from [52].
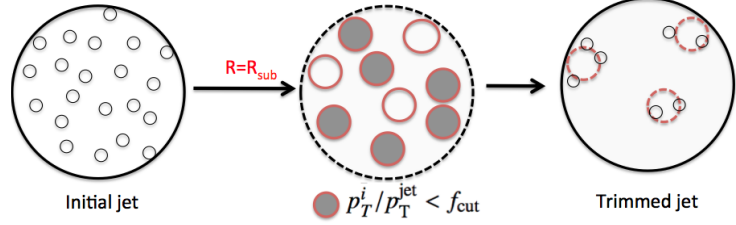
Pruning [54], like Trimming, takes as input a sequentially clustered jet, and then reclusters the jet constituents. At each clustering step $j_1, j_2 \rightarrow j$, labeled such that $p_{\perp j_1} > p_{\perp j_2}$, two conditions are imposed on the $p_\perp$ asymmetry and the constituent distance:

$$z = \frac{p_{\perp j_2}}{p_{\perp j}} < z_{\text{cut}} \tag{1.57a}$$

$$dR_{j_1 j_2} > D_{\text{cut}} \tag{1.57b}$$

where z_{cut} and D_{cut} are free parameters (typically 0.10 and $0.5 \times M_J/p_\perp$, cf. (1.53)). If both conditions are met, the recombination $j_1, j_2 \rightarrow j$ is vetoed, and j_2 is discarded. These conditions are intended to veto recombinations which are thought to be artifacts of the clustering procedure itself rather than any underlying, physical, heavy-particle decay. The procedure is shown in Fig. 1.21.

Figure 1.21: Diagrammatic representation of the Pruning method for jet grooming. Adapted from [52].
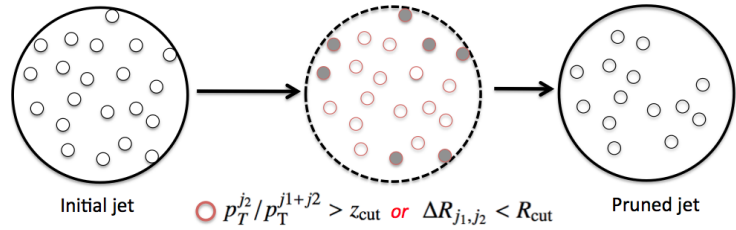
When reconstructing a booted hadronically decaying boson, one must decide on which clustering algorithm and distance parameter to use, where two competing effects must be considered: if the parameter is small, there is a risk of lateral leakage, meaning that wide-angle particles from the jet will fall outside its "radius" and not be recombined; on the other hand, if the parameter is large, contamination from the UE and pile-up will result in too many particles being recombined.

ATLAS Run 1 analyses [14, 15, 16] have found that the Cambridge-Aachen algorithm with $R = 1.2$, abbreviated as $\text{C/A}^{R=1.2}$, is performant and works well with mass-drop filtering for boson substructure tagging.

Pile-up mitigation

The grooming methods presented above are mainly intended to remove the soft contributions from individual hard events. But in the presence of additional simultaneous minbias collisions, cf. Sec. 1.2, each depositing an average energy of $d\langle\rho\rangle/dN_{PU} \approx 0.4\,\text{GeV}$ per unit of $\eta \times \phi$ [55], the reconstructed jet energies and masses will be biased towards larger values by the recombination of uniformly distributed, soft particles from pile-up.

In the presence of large numbers of pile-up collisions, standard grooming methods may be complemented by specialised pile-up mitigation techniques. Below, be briefly describe a few such methods:

Jet area median subtraction [56], or simply 'area subtraction', uses the assumption of uniformity of the pile-up energy deposition to observe that a jet of area A will—on average—acquire an increase in momentum of $A\rho$, where ρ is the energy density of diffuse noise in the event. This method then proposes to remove this contribution using either scalar or four-vector subtraction:

$$p_\perp^{(\text{sub})} = p_\perp - A\rho \tag{1.58a}$$

$$p_\mu^{(\text{sub})} = p_\mu - A_\mu\rho \tag{1.58b}$$

These assignments require a meaningful definition of the jet area, which generally differ significantly from $\pi^2 R$, for jets with radius parameter R. It may be calculated[44] [57] by clustering a jet after adding a number of infinitesimally soft particles distributed randomly in η and ϕ, called "ghosts", to the event:

$$A(\{g_i\}) = \frac{N_g}{\nu_g} \tag{1.59a}$$

$$A_\mu(\{g_i\}) = \frac{1}{\nu_g} \sum_{g_i \in \text{jet}} g_{\mu i} \tag{1.59b}$$

where N_g is the number of ghosts clustered in the jet, ν_g is the number density of ghosts in $\eta \times \phi$, and g_μ is the ghost four-momentum. The area estimate is then found by averaging over several such clusterings, using different sets $\{g_i\}$, in the limit where $\nu_g \to \infty$.

Similarly, the energy density may be found either as the median density in a collection of smaller-R jets or in a collection of grid patches in $\eta - \phi$

$$\rho = \underset{\text{jet } j}{\text{median}} \left[\left\{ \frac{p_{\perp j}}{A_j} \right\} \right] \tag{1.60a}$$

$$\rho = \underset{\text{patch}}{\text{median}} \left[\left\{ \frac{\sum_{i \in \text{patch}} p_{\perp i}}{A_{\text{patch}}} \right\} \right] \tag{1.60b}$$

Neutral proportional to charged [58], or 'NpC', also works as a jet four-vector subtraction method, but attempts to use local correlation between charged and neutral particles, in the hope that our ability to distinguish between tracks from the hard interaction vertex and

[44] An underlying assumption is, that all considered jet clustering algorithms are infrared-safe, which is the case for all the sequential recombination algorithms presented above.

those from pile-up vertices, see above, will allow for an improvement over standard area subtraction. Specifically, NpC takes as input some jet, identifies the charged particle constituents of the jet which are associated to pile-up vertices, and cluster these into a separate jet (chg-PU). The assumption of local correlation then amounts to the hypothesis that the contribution from neutral pile-up particles is proportional to that of the charged particles (thus the method's name) through

$$\gamma_0 \equiv \left\langle \frac{\sum_{i \in \text{charged particles}} p_{\perp i}}{\sum_{i \in \text{all particles}} p_{\perp i}} \right\rangle_{\text{events}} \tag{1.61}$$

such that the subtraction from the jet four-momentum, p_μ, takes the form

$$p_\mu^{(\text{sub})} = p_\mu - \frac{1}{\gamma_0} p_\mu^{(\text{chg-PU})} \tag{1.62}$$

The indications are, however, that this method does not perform better than simple area subtraction.

SoftKiller [59] works on the particle level, prior to jet clustering, by discarding all particles with $p_\perp$ below some threshold $p_{\perp\text{cut}}$, which is determined on an event-by-event basis such as to ensure that the total energy density in the event, computed as in Eq. (1.60b), is zero, i.e.

$$p_{\perp\text{cut}} = \underset{\text{patch}}{\text{median}} \left[\underset{i \in \text{patch}}{\max} \{p_{\perp i}\} \right] \tag{1.63}$$

The remaining particles then represent a cleaned event and may be subjected to jet clustering. This method is motivated by the fact that the main feature discriminating pile-up particles from hard scatter ones is their $p_\perp$, cf. the steep spectrum in Eq. (1.42).

Pile-up per particle identification [60], or 'PuPPI', is also a particle-level algorithm, which works by computing a local shape for each particle in the event, effectively $\alpha = \log \sum_{j \in \text{cone}} p_{\perp j}/dR_{ij}$ in a cone[45] of radius $R_0 \approx 0.3$ around each particle i. These shapes are observed to differ for hard interaction particles and pile-up particles: the former will be accompanied by a large colliear activity due to parton showers; the latter will have weight which are $p_\perp$ suppressed, see above, and the particles within the cone will mainly be due to chance, because of the lack of splitting at low $p_\perp$. Based on the local shape (possibly using track-vertexing information, like NpC), each particle is assigned a weight $w \in [0,1]$ and is rescaled as $p_\mu \to w \times p_\mu$. Particles with $w < w_\text{cut}$ or rescaled $p_\perp < p_{\perp\text{cut}}$ are deemed to be from pile-up and are discarded. As for the SoftKiller method, the remaining (rescaled) particles then constitute a cleaned event, which may be used for subsequent jet clustering.

Of the four methods presented above, area subtraction is by far the most used method [55]. Attempts to improve jet-level methods, like the NpC methods, have not been successful. The particle-level methods, however, hold promise, and in the following these will serve as a main source of inspiration.

[45] Excluding particles within a much smaller cone of radius $R_\text{min} \approx 0.02$ around the particle in order to avoid collienar divergences.

II Wavelets

2.1 Wavelet theory

As they said in *Monty Python's Flying Circus*: "And now for something completely different", yet maybe not quite. In the previous chapter we introduced the basic of particle interactions, the use of particle collisions in searches for new physics, the detection of particles using the ATLAS detector, and the reconstruction of physical particles from detector measurements. We noted that many signs of new physics will be visible in final states with two massive gauge bosons and concluded the chapter with a discussion of their reconstruction as "jets" and the obscuring role played by pile-up collisions. In the following sections, we will develop pile-up mitigation techniques based on *wavelets*, and the present section is devoted to an introduction of these mathematical constructs in and of themselves.

■ Multiresolution analysis

The Fourier transform has proved an indispensable tool in the natural sciences, especially within physics, allowing for the study of frequency or momentum information of functions and for the efficient representation of signals exhibiting periodic structure. Similarly, the Fourier transform allows for the study of signals—in "position space"—at different frequency or, equivalently, resolution scales. This separation by frequency scale is the feature which will be of our main interest. However, the Fourier transform has the limitation that it is global: each frequency component carries no information about its localisation in space/time; information which might be valuable.

One way to address this problem is to introduce the windowed, or short-time, Fourier transform, where the signal is multiplied by a fixed-width windowing function, e.g. a gaussian, prior to the Fourier transform. While this methods does introduce spatial localisation of the various frequency components, the fact that the windowing function has a fixed width means that all frequency components are localised with the same spatial resolution (i.e. the width of the windowing function). However, due to their smaller wavelengths, high-frequency components may be localised to a larger precision than their low-frequency counterparts. Therefore, we will start by developing the fundamental concepts of a multiresolution analysis, i.e. a representation of a function at various levels of resolution, which encodes position-frequency information in an efficient way. The following exposition closely follows[1] [61, 62, 63].

Defining properties

Let $L_2(\mathbb{R})$ be the space of all square integrable fuctions[2] $f : \mathbb{R} \to \mathbb{R}$, with inner product $\langle \cdot, \cdot \rangle$

$$\langle f, g \rangle = \int_{-\infty}^{\infty} f(x)g(x)\, \mathrm{d}x \qquad (2.1)$$

and norm $\| \cdot \|$, $\|f\| \equiv \langle f, f \rangle$, such that

$$L_2(\mathbb{R}) = \{\, f : \mathbb{R} \to \mathbb{R} \mid \|f\| < \infty \,\} \qquad (2.2)$$

[1] Note, that in this presentation we only focus on the orthonormal wavelet bases and thus omit the discussion of over-complete/redundant bases. This has the consequences that the "continuous wavelet transform" mentioned here correspond to the discrete transform in [61]. Therefore, the present distinction between "continuous" and "discrete" only relates to whether the signal being processed is itself continuous (function) or discrete (vector or matrix, in the sense of [62]).

[2] One might equally well consider all function $f : \mathbb{R} \to \mathbb{C}$, in which case a complex conjugation should be introduced in Eq. (2.1).

Suppose, then, that we want a representation of some function $f \in L_2(\mathbb{R})$ at various resolution levels.

In order to develop a multiresolution analysis we need a series of function spaces V_m, for integer m, which are nested as

$$\{0\} \subset \cdots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots \subset L_2(\mathbb{R}) \qquad \text{(P1)}$$

The projection of f onto V_m is denoted by the operator P_m, and correspondingly $P_m f$ is called representation of f at *scale m*. Higher scales mean finer resolution. Since we want to represent all functions $f \in L_2(\mathbb{R})$ we must require that in $\bigcup_{m \in \mathbb{Z}} V_m$ is dense in $L_2(\mathbb{R})$, i.e.

$$\overline{\bigcup_{m \in \mathbb{Z}} V_m} = L_2(\mathbb{R}) \qquad \text{(P2)}$$

or equivalently that $\lim_{m \to \infty} P_m f = f$. Here $\overline{}$ denotes the closure of a set and $\cup$ denotes the union of sets. Similarly, we will have $\lim_{m \to -\infty} V_m = \{0\}$, such that

$$\bigcap_{m \in \mathbb{Z}} V_m = \{0\} \qquad \text{(P3)}$$

where $\cap$ denotes the intersection of sets. The concept of "multiresolution" then arises if we choose to relate the subspaces by a dyadic scaling[3], such that

$$g(x) \in V_m \quad \Longleftrightarrow \quad g(2x) \in V_{m+1} \qquad \text{(P4)}$$

That is, all function spaces V_m are simply scaled versions of a single space, V_0. Finally, we require each space V_m to be spanned by integer translates of the same fundamental basis function ϕ_m, i.e.

$$V_m = \text{span}\{\, \phi_m(x - l) \mid l \in \mathbb{Z} \,\} \qquad \text{(P5)}$$

The properties (P1–5) then serve as the definition for our multiresolution analysis.

Scaling relation

Let $\phi(x)$ be some fixed, normalised function, called the *scaling function*, whose integer translations span V_0. Defining $\phi_{0,l}(x) \equiv \phi(x - l)$ this means

$$\langle \phi_{0,l}, \phi_{0,k} \rangle = \delta_{l,k} \quad \text{and} \quad V_0 = \text{span}\{\, \phi_{0,l}(x) \mid l \in \mathbb{Z} \,\} \qquad \text{(2.4)}$$

Since the spaces V_m are related to V_0 by a simple scaling cf. (P4), we can define

$$V_m \ni \phi_{m,l}(x) \equiv \sqrt{2^m}\, \phi(2^m x - l) \qquad \text{(2.5)}$$

such that

$$\begin{aligned}
\langle \phi_{m,l}(x), \phi_{m,k}(x) \rangle &= 2^m \int \phi_0(2^{-m}x - k)\phi_0(2^{-m}x - l)\, \mathrm{d}x \\
&= 2^m 2^{-m} \int \phi_0(x' - k)\phi_0(x' - l)\, \mathrm{d}x' \\
&= \langle \phi_{0,l}, \phi_{0,k} \rangle = \delta_{l,k} \qquad \text{(2.6)}
\end{aligned}$$

[3] Scaling by powers other that 2 is also possible, but this choice is particularly nice.

and conclude that $\{\,\phi_{m,l}(x) \mid l \in \mathbb{Z}\,\}$ constitutes an orthonormal basis for V_m by (P4). The projection $P_m f$ is then given by

$$P_m f = \sum_{k \in \mathbb{Z}} \langle f, \phi_{m,k} \rangle \, \phi_{m,k} \tag{2.7}$$

But since $V_m \subset V_{m+1}$ we must have

$$\phi_{m,0}(x) = \sum_{k \in \mathbb{Z}} a_k \phi_{m+1,k}(x) \tag{2.8}$$

The (finite) set of coefficient $\{a_k\}$ are called the filter coefficients and must satisfy

$$\sum_{k \in \mathbb{Z}} a_k^2 = 1 \qquad \text{and} \qquad \sum_{k \in \mathbb{Z}} a_k a_{k+2l} = \delta_{0,l} \tag{2.9}$$

ensuring normalisation and orthogonality, resp.

That is, each scaling function at scale m can be expressed as a linear combination of scaling function at the next higher scale $m + 1$. Eq. (2.8) is called the *scaling relation* and highlights the self-similarity, or fractal nature, of scaling functions satisfying (P1–5). Furthermore, Eq. (2.5) also shows that the width of the scaling function goes as 2^{-m}, i.e. changes with frequency. This is exactly the feature missing from the windowed Fourier transform.

Wavelet functions

So far we have only talked about representing a function at varying degrees of granularity, but not mentioned the frequency aspect. To obtain a frequency representation, we note that (P1) states that the function spaces V_m are, quite naturally, redundant. However, we may be interested in the *loss of information* in going from scale $m + 1$ to scale m, i.e. the frequency information stored at scale $m + 1$. This will allow us to study the information stored in different frequency bands, labeled by the scale index m.

Therefore, we define W_m to be the orthogonal complement of V_m in V_{m+1}:

$$V_{m+1} = V_m \oplus W_m \qquad \text{and} \qquad V_m \perp W_m \tag{2.10}$$

where $\oplus$ is the direct sum. It can then be shown[4] that

$$\overline{\bigoplus_{m \in \mathbb{Z}} W_m} = L_2(\mathbb{R}) \tag{2.11}$$

Analogously to the case of the function spaces V_m, we are looking for *wavelet functions*[5] $\psi : \mathbb{R} \to \mathbb{R}$ such that $\{\,\psi(x - l) \mid l \in \mathbb{Z}\,\}$ forms a basis for W_0 and let the basis functions for W_m be given by the equivalent of Eq. (2.5)

$$\psi_{m,l}(x) = \sqrt{2^m} \, \psi\left(2^m x - l\right), \qquad \psi_{0,0}(x) \equiv \psi(x) \tag{2.12}$$

Since $W_m \subset V_{m+1}$, the wavelet function ψ also satisfies an equation similar to the scaling relation in Eq. (2.8)

$$\psi_{m,0}(x) = \sum_{k \in \mathbb{Z}} b_k \phi_{m+1,k}(x) \tag{2.13}$$

[4] To show that $\bigoplus_{m \in \mathbb{Z}} W_m$ is dense in $L_2(\mathbb{R})$, we note that $\bigcup_{m \in \mathbb{Z}} V_m$ is dense in $L_2(\mathbb{R})$ and that $V_m \subset V_n$ for $m < n$. Then V_m is dense in $L_2(\mathbb{R})$ for $m \to \infty$, and, from Eq. (2.10), $V_m = V_M \oplus \left(\bigoplus_{m' \in [M,m[} W_{m'} \right)$. Since $\lim_{M \to -\infty} V_M = \{0\}$ and $\{0\} \in W_m$ we have that $\bigoplus_{m' \in [M,m[} W_{m'}$ is dense in $L_2(\mathbb{R})$ for $m \to \infty$ and $M \to -\infty$. We conclude that $\bigoplus_{m \in \mathbb{Z}} W_m$ is dense in $L_2(\mathbb{R})$.

[5] The existence of such functions is proved in detail in [61].

where the coefficients b_k may be expressed in terms of the filter coefficients as

$$b_k = (-1)^k a_{N-1-k} \qquad \text{where} \qquad N = |\{ a_k \}| \qquad (2.14)$$

Since the wavelet functions $\psi_{m,l}$ span W_m, and since the direct sum of the W_m spaces is dense in $L_2(\mathbb{R})$ cf. Eq. (2.11), we see that any function may be expressed in terms of the wavelet functions *alone*, about which we know that: they encode frequency information (through the scale-label m); they encode position information (through the position-label l, cf. Eq. (2.12)); their width scales with the resolution; and they exhibit the same self-similar behaviour as the scaling function, through the wavelet-space equivalent of (P4). Finally, we note that the multiresolution analysis is completely specified by the set $\{a_k\}$.

Example wavelet dictionaries

So far the treatment of the multiresolution analysis has been completely abstract. Above we presented a general, desirable prescription for a multiresolution analysis, but as yet it is not yet clear whether scaling functions satisfying (P1–5) actually exist. We now give a few simple examples of actual wavelet dictionaries, i.e. the fundamental set $\{ \phi(x), \psi(x) \}$ from which the complete set of scaling and wavelet basis functions $\{ \phi_{m,l}(x) \mid m,l \in \mathbb{Z} \} \cup \{ \psi_{m,l}(x) \mid m,l \in \mathbb{Z} \}$ is specified.

Possibly simplest space in which to represent some function at varying degrees of resolution is the space of piecewise constant functions, i.e.

$$V_m = \left\{ f \in L_2(\mathbb{R}) : f(x) = \begin{cases} \text{const.} & x \in [2^m l,\, 2^m(l+1)[\\ 0 & \text{otherwise} \end{cases} \, \middle| \, l \in \mathbb{Z} \right\}$$
$$(2.15)$$

These space are exactly those spanned by the "top-hat" function

$$\phi(x) = \begin{cases} 1 & x \in [0,\, 1[\\ 0 & \text{otherwise} \end{cases} \qquad (2.16)$$

such that the $\phi_{m,l}$ given by Eq. (2.5) form an orthonormal basis. From the scaling equation Eq. (2.8) we find that the filter coefficients for the scaling function are

$$a_0 = a_1 = \frac{1}{\sqrt{2}}; \qquad |\{ a_k \}| = 2 \qquad (2.17)$$

Knowing the filtering equations we can directly construct the coefficients $\{ b_k \}$ for the fundamental wavelet function corresponding to Eq. (2.16), namely

$$b_0 = (-1)^0 a_{2-1-0} = \quad a_1 = \quad \frac{1}{\sqrt{2}} \qquad (2.18\text{a})$$

$$b_1 = (-1)^1 a_{2-1-1} = -a_0 = -\frac{1}{\sqrt{2}} \qquad (2.18\text{b})$$

giving the wavelet function

$$\psi(x) = \begin{cases} 1 & x \in [0, 1/2[\\ -1 & x \in [1/2, 1[\\ 0 & \text{otherwise} \end{cases} \tag{2.19}$$

with $\psi_{m,l}$ given by Eq. (2.12). This is the the Haar wavelet [64] which is the standard example of a one dimensional, continuous wavelet. The Haar scaling and wavelet functions are shown in Fig. 2.1(a).

More generally, it is possible to construct a family of orthonormal wavelets with compact support by requiring that wavelet functions with $|\{a_k\}| = N$ have $N/2$ vanishing moments. This family is called the Daubechies wavelets (D_N) and the Haar wavelet is the special case for $N = 2$. Finding the filter coefficients for this family is laborious, but as an example the coefficients are shown for D_2, D_4, and D_8 in Tab. 2.1.

Similarly, constructing the scaling and wavelet functions is involved and only a few orthonormal wavelet bases have a closed form expression. For instance, the D_N functions can only be constructed recursively beyond $N = 2$ (to an arbitrary precision), using the scaling relation in Eq. (2.8). The functional form D_4 and D_8 are shown in Figs. 2.1(b) and (c).

| k | a_k | | |
	D_2	D_4	D_8
0	$1/\sqrt{2}$	0.4830	0.2304
1	$1/\sqrt{2}$	0.8365	0.7148
2		0.2241	0.6309
3		−0.1294	−0.0280
4			−0.1870
5			0.0308
6			0.0329
7			−0.0106

Table 2.1: Filter coefficients for three members of the Daubechies wavelet family D_N. D_2 is identical to the Haar wavelet.

(a) Haar (D_2).

(b) Daubechies 4 (D_4).

(c) Daubechies 8 (D_8).
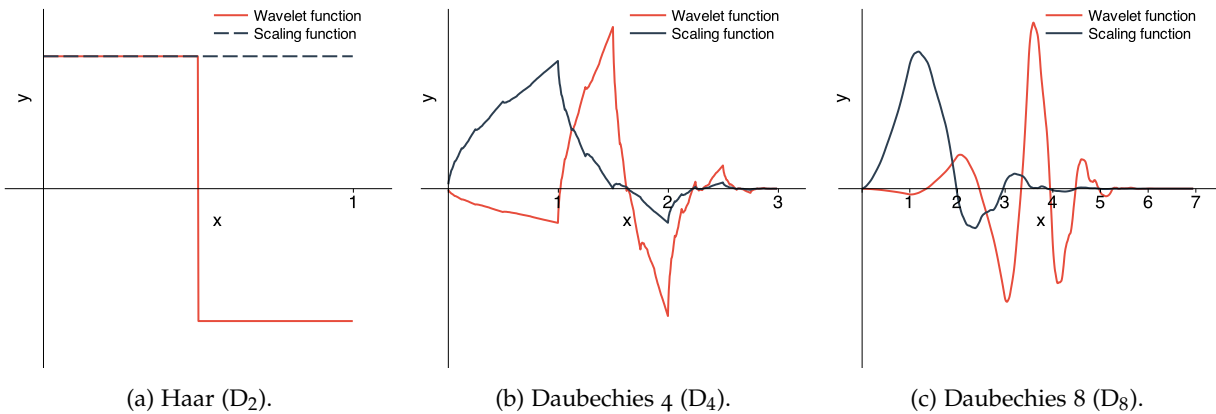
Figure 2.1: Three examples of wavelet dictionaries. Units on the y-axis are arbitrary.

Here it is seen how increasing the number of filtering coefficients (the *degree* of the wavelet basis) leads to a better frequency-localisation (i.e. a clearer oscillatory behaviour—with an increasing number of "taps"—and thus a faster-decaying Fourier transform) at the expense of poorer spatial localisation, since the scaling and wavelet functions for D_N both have support $[0, N-1[$.

■ Continuous wavelet transform

Above we have discussed the concept of a multiresolution analysis and the role played by the scaling and wavelet functions. Now we turn to the actual wavelet transform. The wavelet transform proceeds in two directions, like e.g. the Fourier transform: the forward transform (analysis), i.e. the representation of a signal in the space of wavelet

coefficients, and the inverse transform (synthesis), i.e. the reconstruction of a signal from a set of wavelet coefficients. In the following we distinguish between a continuous wavelet transform (CWT), where the input signal is inherently continuous (e.g. a function), and a discrete transform (DWT), where the input signal is inherently discrete (e.g. a vector or a matrix).

Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be some function, called the "signal", to which we want to apply a continuous wavelet transform. Let 0 and J be the lowest and highest scales (frequencies) at which we are interested in representing the signal[6], which is assumed to be scaled to the range $[0, 1]$. In a physics context, these are typically given/limited by the physical extent of the signal and the experimental temporal/spatial-resolution, respectively. We choose some wavelet dictionary $\{ \phi_{m,l} \mid m \in [0, J], l \in \mathbb{Z} \} \cup \{ \psi_{m,l} \mid m \in [0, J - 1], l \in \mathbb{Z} \}$, where in general $l \in \mathbb{Z}$, but in practice it will be restricted to $[0, 2^m - 1]$, due to the dyadic nature of the wavelet dictionary, since all other values will result in the function being outside the range of the signal, and all related coefficients will therefore be identically zero. We will use this restricted range below. The continuous wavelet transform then proceeds as follows.

Analysis

The projection at scale m of the signal f onto the space V_m, $P_m f$, can be found from the projection at the next higher scale, $P_{m+1} f$, by the prescription in Eq. (2.7)

$$P_m f = \sum_{l=0}^{2^m - 1} c_{m,n} \, \phi_{m,l} \qquad \text{with} \quad c_{m,n} = \langle P_{m+1} f, \phi_{m,l} \rangle \qquad (2.20)$$

This acts as a *low-pass filter*, by effectively averaging out the details at frequency scale $m + 1$. This means that we can find projections at coarser resolutions/lower frequencies by successively applying the above prescription. The iterative process then starts from the highest chosen scale, by

$$P_J f = \sum_{l=0}^{2^J - 1} c_{J,l} \, \phi_{J,l} \qquad \text{with} \quad c_{J,l} = \langle f, \phi_{J,l} \rangle \qquad (2.21)$$

Orthogonally to the low-pass filter, at each scale m we also apply a *high-pass filter*, to pick up the high-frequency information otherwise lost by the projection to the next lower scale. Therefore, the *detail at scale m, Q_m*, lost in P_m, is given by the projection under $L_2(\mathbb{R})$-inner product of P_{m+1} onto the space W_m spanned by the wavelet functions $\{ \psi_{m,l} \mid l \in [0, 2^m - 1] \}$

$$Q_m f = \sum_{l=0}^{2^m - 1} d_{m,l} \, \psi_{m,l} \qquad \text{with} \quad d_{m,l} = \langle P_{m+1} f, \psi_{m,l} \rangle \qquad (2.22)$$

such that, by Eq. (2.10)

$$Q_m f = P_{m+1} f - P_m f \qquad (2.23)$$

[6] This choice of scales may accommodate physical frequencies tending towards $\mp \infty$ by a rescaling of the signal. We will generally assume that the signal has finite support.

After having applied the high-pass filter and found the detail co-efficients at scale $J-1$, $d_{J-1,l}$, another low-pass filter is applied to P_J, and so forth. This means that we successively decompose the signal into contributions from a set of orthogonal function-spaces W_m, each encoding the frequency information at scale m, the direct sum of which is (sufficiently, for $J < \infty$) dense in $L_2(\mathbb{R})$, cf. Eq. (2.11). In this way we separate the contributions from the signal coming from different frequency scales. The iterative process ends at $P_0 f$ (the average of the signal wrt. the choice of scaling function).

Looking at Fig. 2.1(a), the action of the low- and high-pass filters is clear: the Haar scaling function ϕ computes the average of the signal on its support, while the wavelet function ϕ computes the difference lost in the average. The effect is the same for $D_{N>2}$, but less clear.

Synthesis

Since the detail coefficients $d_{m,l}$ encode the information lost under the operation $P_{m+1}f \to P_m f$, we can, cf. Eq. (2.23), from the projection at some scale m, reconstruct the projection at the next higher scale, as

$$P_{m+1}f = P_m f + \sum_{l=0}^{2^m-1} d_{m,l}\, \psi_{m,l} \qquad (2.24)$$

This means that the set of *wavelet coefficients* $\{\, c_{0,0} \,\} \cup \{\, d_{m,l} \mid m \in [0, J-1],\ l \in [0, 2^m-1]\ \}$ thus completely specifies the signal f, up to frequency scale J, and therefore the signal may be represented as

$$f \approx \hat{f} = P_J f = P_m f + Q_m f = P_0 f + Q_0 f + \cdots + Q_m f$$

$$= c_{0,0}\,\phi_{0,0} + \sum_{m=0}^{J-1} \sum_{l=0}^{2^m-1} d_{m,l}\, \psi_{ml} \qquad (2.25)$$

■ Discrete wavelet transform

In the cases which will be of interest to us (see Sec. 2.2), the input signal will be discrete rather than continuous. We therefore need to develop methods specific to this situation, although the nomenclature and general structure of the wavelet transform carries over from the continuous case.

In the discrete case, we do not *per se* have a set of pre-defined basis functions, with which we can take the inner product of the signal in order to get the low- and high-pass coefficients $c_{m,l}$ and $d_{m,l}$, see Eqs. (2.20) and (2.22). Instead, we need to use the filter coefficients directly. One way to implement the discrete wavelet transform in this way is through matrix algebra.

One dimension

Assume that we have a one-dimensional, discretised signal $\mathbf{f}$, given as a column matrix of length $|\mathbf{f}| = 2^J$ for some $J \in \mathbb{N}$:

$$\mathbf{f} = \begin{bmatrix} \mathbf{f}[0] \\ \mathbf{f}[1] \\ \vdots \\ \mathbf{f}[2^J - 2] \\ \mathbf{f}[2^J - 1] \end{bmatrix} \tag{2.26}$$

This requirement on the length of the signal is necessary due to the dyadic scaling in (P4). The low-pass filter applied to $\mathbf{f}$, $L\mathbf{f}$ going from scale $J \to J - 1$, is then implemented as the explicit convolution of the entries of $\mathbf{f}$ with the filter coefficients $a[k] \equiv a_k$, where $N = |\{a[k]\}|$

$$L\mathbf{f}[n] = \sum_{k=0}^{2^J-1} a[k]\mathbf{f}[2n + N/2 - k] \qquad \text{with} \quad n \in [0, 2^{J-1} - 1] \quad (2.27)$$

assuming periodicity of $\mathbf{f}$, such that $\mathbf{f}[-1] = \mathbf{f}[2^J - 1]$, etc. Here the $N/2$ term makes sure that the convolution is centered on "entry" $2n + 1/2$ in $\mathbf{f}$. Note, that this means that for $N > 2$ the requirement of periodicity means that for the smallest and largest n, the convolution will pick up contributions from opposite ends of $\mathbf{f}$ which may be problematic.

The convolution yielding each entry n in $L\mathbf{f}$ can be seen as a matrix inner product of $\mathbf{f}$ with a row matrix of the form

$$\begin{bmatrix} \cdots & 0 & a[N-1] & \cdots & a[1] & a[0] & 0 \cdots \end{bmatrix} \tag{2.28}$$

Since this is true for each entry, the full low-pass filter may be represented as a $(2^{J-1} \times 2^J) \cdot (2^J \times 1)$ matrix inner product

$$L\mathbf{f} = \mathbf{L}_{J-1} \cdot \mathbf{f} \tag{2.29}$$

where, for each low-pass operation, we have

$$\mathbf{L}_m = \left.\begin{bmatrix} \ddots & & \ddots & & \ddots & & \ddots & & & & \\ \cdots & a[N-1] & \cdots & & a[1] & a[0] & 0 & & 0 & 0 & 0 & \cdots \\ \cdots & 0 & 0 & a[N-1] & \cdots & & a[1] & a[0] & 0 & 0 & \cdots \\ \cdots & 0 & 0 & 0 & 0 & a[N-1] & \cdots & a[1] & a[0] & \cdots \\ & & & & & & \ddots & \ddots & \ddots & \ddots \end{bmatrix}\right\} 2^m \tag{2.30}$$

$$\underbrace{}_{2^{m+1}}$$

Similarly, the high-pass filter for $J \to J - 1$ may be written as

$$H\mathbf{f} = \mathbf{H}_{J-1} \cdot \mathbf{f} \tag{2.31}$$

where we similarly have

$$\mathbf{H}_m = \left. \begin{bmatrix} \ddots & & \ddots & & \ddots & & \ddots & & & & \\ \cdots & b[N-1] & \cdots & b[1] & b[0] & 0 & & 0 & 0 & 0 & \cdots \\ \cdots & 0 & 0 & b[N-1] & \cdots & b[1] & b[0] & 0 & 0 & \cdots \\ \cdots & 0 & 0 & 0 & 0 & b[N-1] & \cdots & b[1] & b[0] & \cdots \\ & & & & & \ddots & \ddots & \ddots & \ddots & \end{bmatrix} \right\} 2^m \qquad (2.32)$$

$$\underbrace{}_{2^{m+1}}$$

We then see that

$$P_m \mathbf{f} \equiv L^{J-m} \mathbf{f} = \mathbf{L}_m \mathbf{L}_{m+1} \cdots \mathbf{L}_{J-1} \mathbf{f} = \mathbf{L}_m P_{m+1}; \qquad P_J \equiv \mathbf{f} \qquad (2.33)$$

corresponds to the projection $P_m f$ at scale m in the continuous case. Similarly, we see that $\mathbf{H}_m$ picks out the detail coefficients at level m, such that

$$Q_m \mathbf{f} \equiv \mathbf{H}_m P_{m+1} \mathbf{f} = \begin{bmatrix} d_{m,0} & d_{m,1} & \cdots & d_{m,2^m-1} \end{bmatrix} \qquad (2.34)$$

This means that the iterative nature of the forward wavelet transform in the continuous case carries over, such that we apply low-pass filters to go down in resolution in steps of $m \to m-1$, at each step using the high-pass filter to pick up the frequency information at scale m. The result is a set of wavelet coefficients similar to those in Eq. (2.25), which may be stored in a column vector

$$\mathbf{w} = \left[\overbrace{c_{0,0}}^{P_0 f} \ \overbrace{d_{0,0}}^{Q_0 f} \ \overbrace{d_{1,0} \ d_{1,1}}^{Q_1 f} \ \cdots \ \overbrace{d_{J-1,0} \ \cdots \ d_{J-1,2^{J-1}-1}}^{Q_{J-1} f} \right]^\top$$
$$\underbrace{\phantom{c_{0,0}}}_{1} \ \underbrace{\phantom{d_{0,0}}}_{1} \ \underbrace{\phantom{d_{1,0} \ d_{1,1}}}_{2} \qquad \underbrace{\phantom{d_{J-1,0} \cdots d_{J-1}}}_{2^{J-1}}$$
$$(2.35)$$

The number of coefficients is exactly

$$1 + \sum_{m=0}^{J-1} 2^m = 2^J \qquad (2.36)$$

as required for an exact representation of $\mathbf{f}$.

Similarly to Eq. (2.23), we note that

$$P_{m+1} \mathbf{f} = \mathbf{L}_m^{-1} P_m \mathbf{f} + \mathbf{H}_m^{-1} Q_m \mathbf{f} \qquad (2.37)$$

where $\mathbf{L}_m^{-1}$ and $\mathbf{H}_m^{-1}$ are the inverse low- and high-pass filter matrices at scale m. Recalling that we are only considering orthonormal bases, we have[7]

$$\mathbf{L}_m \mathbf{L}_m^\top = \mathbf{I}_{m \times m}$$

where $\mathbf{I}_{m \times m}$ is the $m \times m$ identity matrix, and similarly for $\mathbf{H}_m$, meaning that the inverses of the low- and high-pass filter matrices are just their transposes.

From this set of wavelet coefficients, it is then possible to reconstruct the original signal using a synthesis similar to that for the

---

[7] Since we have

$$(\mathbf{L}_m)_{r,c} = a[2r + N/2 - c] \qquad (2.38)$$

with rows $r \in [0, 2^m - 1]$ and columns $c \in [0, 2^{m+1} - 1]$, then we can write, using Eq. (2.9)

$$(\mathbf{L}_m \mathbf{L}_m^\top)_{r,c} = \sum_{s=0}^{2^{m+1}-1} (\mathbf{L}_m)_{r,s} (\mathbf{L}_m^\top)_{s,c}$$

$$= \sum_{s=0}^{2^{m+1}-1} a[2r + N/2 - s]$$
$$\times a[2c + N/2 - s]$$

$$= \sum_{s=0}^{2^{m+1}-1} a[s]^2 \delta_{0,r-c}$$

$$= \delta_{r,c} \underbrace{\left[\sum_{s=0}^{2^{m+1}-1} a[s]^2 \right]}_{1}$$

which means that

$$\mathbf{L}_m \mathbf{L}_m^\top = \mathbf{I}_{m \times m}$$

continuous transform, cf. Eq . (2.25)

$$\mathbf{f} = P_J\mathbf{f} = \mathbf{L}_{J-1}^\top P_{J-1}\mathbf{f} + \mathbf{H}_{J-1}^\top Q_{J-1}\mathbf{f}$$

$$= \mathbf{L}_{J-1}^\top \left[\mathbf{L}_{J-2}^\top P_{J-2}\mathbf{f} + \mathbf{H}_{J-2}^\top Q_{J-2}\mathbf{f}\right] + \mathbf{H}_{J-1}^\top Q_{J-1}\mathbf{f}$$

$$\vdots$$

$$= \prod_{m=J-1}^{0} \mathbf{L}_m^\top P_0\mathbf{f} + \sum_{i=0}^{J-1}\left(\prod_{m=J-1}^{i+1} \mathbf{L}_m^\top\right)\mathbf{H}_i^\top Q_i\mathbf{f}$$

$$= \prod_{m=J-1}^{0} \mathbf{L}_m^\top \mathbf{c}_0 + \sum_{i=0}^{J-1}\left(\prod_{m=J-1}^{i+1} \mathbf{L}_m^\top\right)\mathbf{H}_i^\top \mathbf{d}_i \qquad (2.39)$$

where $\mathbf{c}_0, \mathbf{d}_m$ are the vectors of wavelet coefficients contained in $\mathbf{w}$.

The "basis vectors" for the discrete wavelet transform are not given directly, as noted above, but found by performing the inverse transform of $\mathbf{w}$ with all but one wavelet coefficients set to 0. The basis function $\phi_{2,2}$ for D$_4$ found by setting $d_{2,2}$ to 1 and performing the inverse transform is plotted in Fig. 2.2 for $|\mathbf{w}| = 2^4, 2^6$, and 2^8.

These basis vectors are orthonormal under standard vector inner product and we see that in the limit of infinite signal size ($J \to \infty$), they correspond to the continuous wavelet functions for the chosen basis, cf. Fig. 2.1(b).

Two dimensions

In some cases (which will be of interest later) the input data is two dimensional[8]: e.g. a matrix, an image, or, of particular interest to us, a detector event display. Therefore, we need to define a prescription for how to deal with these input objects.

Assuming the input to be on the form of an $N \times M$ matrix with $N = 2^J$ and $M = 2^K$ for $J, K \in \mathbb{N}$, one way to define a two dimensional transform[9] is to *first* apply a full forward one-dimensional DWT to each row in the input matrix and to *then* perform another full forward one-dimensional DWT on the resulting rows (i.e. on the wavelet coefficients of the first transform). The resulting $N \times M$ matrix then holds the wavelet coefficients for the two-dimensional forward DWT defined in this way.

For the 1D DWT above, we saw that each of the wavelet coefficients corresponded to an orthonormal "basis vector", but it is not immediately clear than the same holds for the 2D transform. In order to asses this, we may, as above, set all but one of the $N \times M$ matrix of wavelet coefficients to zero, and perform the inverse 2D DWT transform, defined analogously to the forward transform: first, perform an inverse 1D DWT on the columns of the matrix of wavelet coefficients, and the perform another inverse 1D DWT on the rows of the resulting matrix to retrieve the signal.

By doing this we find that each entry in the $N \times M$ matrix of wavelet coefficients *does* indeed correspond to an orthonormal[10] "basis matrix", with *two sets* of scale and position labels, similar to m and l in the one-dimensional case $\psi_{m,l}$,

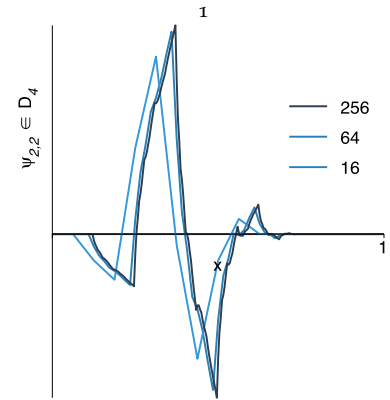Figure 2.2: Examples of a discrete D$_4$ wavelet basis functions ($\psi_{2,2}$) for three different signal lengths. Signal is normalised to length 1.

[8] We will only consider the discrete case.

[9] Which is also the way in which it is implemented in GSL [65], which we will be using.

[10] Under Frobenius matrix inner product, which is the discrete equivalent of the standard inner product on $L_2(\mathbb{R}^2)$.

$$
\mathbf{W} = \begin{bmatrix}
C_{0,0;0,0} & C_{0,0;1,0} & \cdots & C_{0,0;M-1,2^{M-2}-1} \\
C_{1,0;0,0} & C_{1,0;1,0} & \cdots & C_{1,0;M-1,2^{M-2}-1} \\
C_{2,0;0,0} & C_{2,0;1,0} & \cdots & C_{2,0;M-1,2^{M-2}-1} \\
C_{2,1;0,0} & C_{2,1;1,0} & \cdots & C_{2,1;M-1,2^{M-2}-1} \\
C_{3,0;0,0} & C_{3,0;1,0} & \cdots & C_{3,0;M-1,2^{M-2}-1} \\
\vdots & \vdots & \ddots & \vdots \\
C_{N-1,2^{N-2}-1;0,0} & C_{N-1,2^{N-2}-1;1,0} & \cdots & C_{N-1,2^{N-2}-1;M-1,2^{M-2}-1}
\end{bmatrix} \tag{2.40}
$$

where we have dropped the distinction between the *c* and *d* coefficients. The 2D wavelet basis matrices, $\psi_{m_x,l_x;m_y,l_y}$, are simply the *vector outer product* of the corresponding 1D basis vectors, i.e.

$$
\psi_{m_x,l_x;m_y,l_y} = \psi_{m_x-1,l_x}\psi_{m_y-1,l_y}^\top; \qquad m_{x,y} \geq 1 \tag{2.41a}
$$

$$
\psi_{0,0;m_y,l_y} = \phi_{0,0} \quad \psi_{m_y-1,l_y}^\top; \qquad m_y \geq 1 \tag{2.41b}
$$

$$
\psi_{m_x,l_x;0,0} = \psi_{m_x-1,l_x}\phi_{0,0}^\top; \qquad m_x \geq 1 \tag{2.41c}
$$

where $\psi_{m,l}, \phi_{m,l}$ are envisioned as column matrices. The shift in scale indices is made to avoid ambiguities. An example of these 2D basis functions are shown in Fig. 2.3 for the Haar wavelet, on an 8×8 input matrix, map to the $y - \phi$ plane of a detector, such that $m_{x,y} \to m_{y,\phi}$ and $l_{x,y} \to l_{y,\phi}$ (see Sec. 2.2 and onwards for details).

Figure 2.3: The 2D Haar wavelet basis functions in the case of an 8×8 pixel array, each box representing a detector display. Scale $(m_{y,\phi})$ and translation $(l_{y,\phi})$ indices are specified for each function.
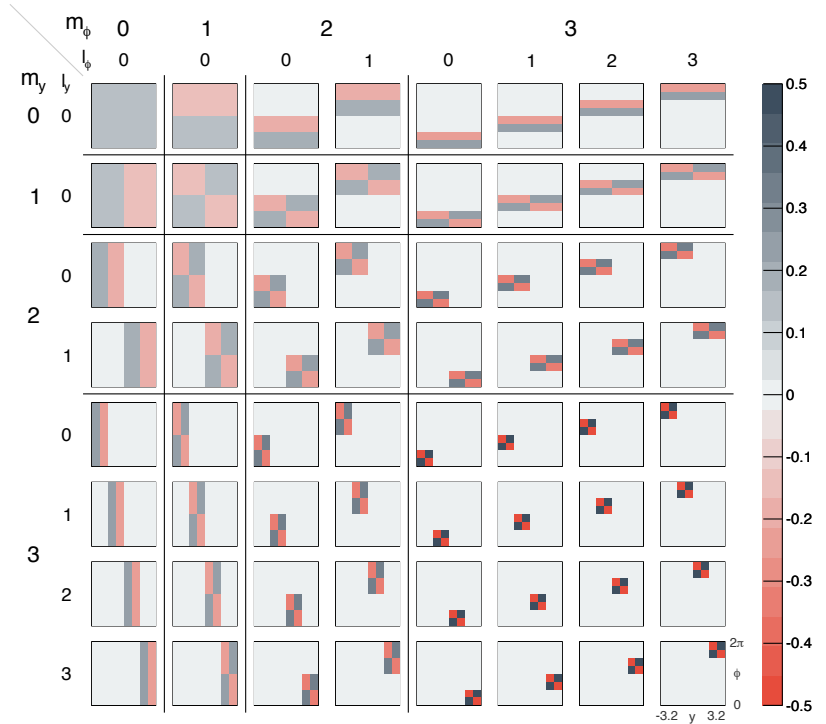
This means that we have developed a prescription for representing a discrete 2D input signal in terms of wavelet basis functions, or "matrices" which encode position and frequency information in the efficient way presented above. This means that we are now equipped to perform a 2D multiresolution analysis on physics-motivated input.

2.2 Wavelets in high energy physics

In the previous chapter we established a theoretical knowledge about wavelets, showed that they have good space- *and* frequency-localisation, and noted that they can be used to efficiently study two-dimensional data at different resolution/frequency scales. Now we need to put our knowledge to good use, and describe how wavelets fit into a particle physics picture. We will start by explaining why we believe that wavelet based methods might be of use in physics analyses, and propose a vocabulary which can hopefully serve as a starting point for a coherent discussion about wavelets in high energy physics.

■ The problem of pile-up

At the time of writing, the LHC is beginning its major second data-taking period (Run 2), where it will be operating at center of mass energies of up to $\sqrt{s} = 14\,\text{TeV}$. Going into Run 2, and Run 3 after that, discussions are ongoing as to what the future of the accelerator and its experiments should be. One option is that being studied by the High-Luminosity LHC (HL-LHC) Project, which is looking into the possibility of operating the LHC at 10 times the current (nominal) luminosity. Such extreme operating conditions will lead to a series of complications for physics analyses, in particular an increase in the number of simultaneous proton collisions at each bunch crossing: pile-up. At the peak instantaneous luminosity of $\sim 7 \times 10^{33}\,\text{cm}^{-2}\,\text{s}^{-1}$, see Fig. 2.4(a), recorded during the 2012 data-taking, the average number of pile-up collisions accompanying any hard scatter interaction was $\langle \mu \rangle = 20.7$, cf. Fig. 2.4(b).
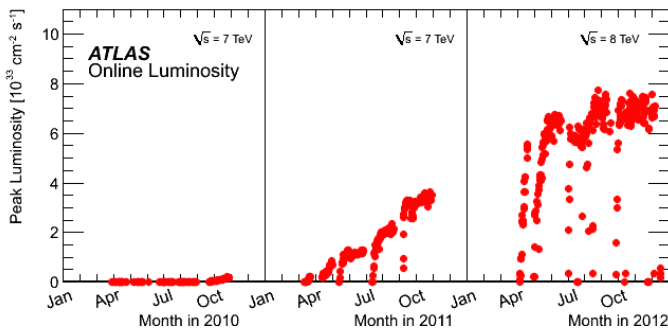
(a) Peak instantaneous luminosity.
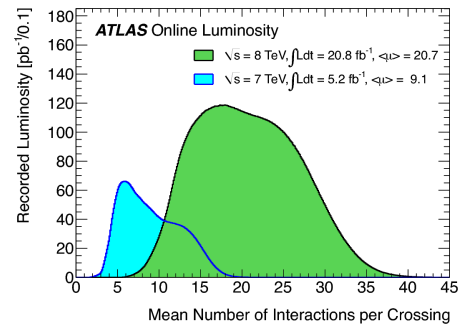
(b) Number of simultaneous interactions per bunch crossing, μ.

The expected average pile-up multiplicity is given by [55]

$$\langle \mu \rangle = \frac{\mathcal{L} \times \sigma_{\text{inel.}}}{N_{\text{bunch}} \times f_{\text{LHC}}} \tag{2.42}$$

where $\mathcal{L}$ is the instantaneous luminosity, $\sigma_{\text{inel.}} \approx 2/3\sigma_{\text{tot}}$ is the total inelastic cross-section, and $N_{\text{bunch}} \times f_{\text{LHC}}$ is average bunch-crossing frequency[11]. Here it is seen, that increasing the luminosity ten-fold—even while keeping the bunch-crossing frequency the same—leads to the possibility of reaching average pile-up multiplicities $\langle \mu \rangle$ upwards of 200. Seeing as each pile-up interactions (being of a soft QCD nature,

Figure 2.4: ATLAS luminosity performance measurements performed during Run 1. From https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults.

[11] Nominally $(25\,\text{ns})^{-1}$, but $(50\,\text{ns})^{-1}$ during Run 1

see Fig. 2.5) on average will deposit roughly $0.4\,\text{GeV}$ of energy per unit $\eta \times \phi$ in the detector, see. Sec. 1.4, the problematic implications for e.g. jet physics become evident: a jet clustered with area A will on average pick up an energy contribution from pile-up equal to $\Delta E_{\text{PU}} \approx N_{\text{PU}} \times 0.4\,\text{GeV} \times A$, which, for an $R = 1.2$ jet at $\langle \mu \rangle = 200$, would amount to an increase in the reconstructed energy of $\gtrsim 300\,\text{GeV}$. Since we are be particularly interested in looking for jets from hadronically decaying W (or Z or H) bosons, each having pole masses of $\mathcal{O}(100\,\text{GeV})$, we see that the spurious contributions from pile-up become dominant. We need to be able to handle the impact of soft pile-up activity in an effective and robust way, if we are ever to perform jet studies with just some degree of precision under these extreme operating conditions. This is where wavelets might be of service.

■ Why wavelets?

A reasonable first question might then be: "why would we want to use wavelets for event cleaning in the first place?". In Sec.1.4 we saw that several motivated, well-tested, and performant methods for pile-up mitigation and jet cleaning already exist. Since wavelets as a tool in high energy physics is quite unconventional, and it deserves some justification before moving forward.

Pragmatist's reason

The pragmatic reason for using wavelets is also motivated by looking at other fields of research. In particular, wavelet methods are used in areas like image manipulation and signal processing [62], where it allows for efficient representation and compression of data, e.g. in the JPEG-2000 image format. Similarly, wavelets are used in other areas of physics for removal of noise and extraction of signal characteristics, e.g. for studying local structures in, and subtracting foregrounds from, Cosmic Microwave Background maps [66]. These qualities, being able to remove unwanted noise, are exactly what we're looking for in the context of pile-up mitigation. Since we are able to cast a hadron collision event in a form similar to an ordinary (monochromatic) image, i.e. in the form of a pixelised event display representing energy deposits in the calorimeter, we might reasonably assume that the benefits of the methods are transferable as well.

Theorist's reason

We can also try to justify the use of wavelets from first principles rather than look to other fields for inspiration. To this end, we can consider two distinct sources of soft particles, which might contaminate our measurement of the hard scatter interaction itself: pile-up and the underlying event (UE).

Recall from Sec. 1.2, that all multi-purpose Monte Carlo generators used at current day HEP experiments use parton showers to simulate
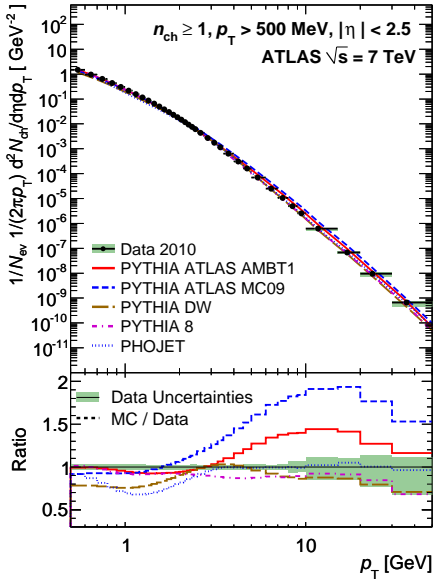

Figure 2.5: Charged particle multiplicity as a function of $p_\perp$ as measured by the ATLAS detector at $\sqrt{s} = 7\,\text{TeV}$. From [28].

the effects of parton radiation beyond the fixed-order of the matrix element. These parton showers—both in the initial and final states—are evolved in some ordering variable Q^2, see Eq. (1.41), from the scale of the hard interaction down to $\mathcal{O}(\Lambda_{\text{QCD}})$, where partons no longer branch off due to colour confinement setting in.

This variable can be chosen, as we have seen, as the transverse momentum p_T^2 or mass M^2 (SHERPA, PYTHIA) of the showering parton, or the splitting angle θ^2 (HERWIG) of the branching parton. This results in branchings which are successively ordered in some angular or energetic manner; an ordering which might be exploited to separate the contribution from the initial, hard scatter from the contributions from soft/wide angle showered partons. Of course, this strict ordering is an artifact of the way event generators are written, but since their end results all describe reality to greater or lesser extent, we can have some confidence that this description of parton showering is not completely off. By using a representation (wavelets) which encodes the angular or frequency information in the event, one might be able to perform this separation in an efficient and elegant way.

Note, that we do not set out to measure the "actual" splitting kernels of QCD[12], cf. Eq. (1.40). Rather, we merely suggest that phenomena, like parton showering, which can be *generated* reliably using an angular ordering, could probably be *decomposed* efficiently using a similar ordering.

The above arguments and parton or jet showering hold both the soft activity coming from the UE as well as that from pile-up. The showering process *per se* is the same for the two ($q \rightarrow qg$ and $g \rightarrow q\bar{q}$ splitting) but the starting points (Q_0^2) are different and so they will result in different radiation patterns, which can (hopefully) be used for separating hard processes from soft contamination in general. This was also the idea behind the PuPPI pile-up mitigation method, cf. Sec. 1.4. Additionally, for pile-up, we have tracking information that allows us to efficiently distinguish between tracks coming from the hard scatter vertex and those coming from secondary interactions. This (auxiliary) information may be used in a wavelet method to improve the effectiveness against pile-up. However, we don't have this option for neutral particles, so all methods for pile-up mitigation are forced to make some assumptions regarding the behaviour of neutral particles, see e.g. area subtraction and NpC in Sec. 1.4.

It is clear, that wavelets in high energy physics in the form presented here, suggested by [67], has large influences from e.g. PuPPI and NpC, which will only become more pronounced as we proceed. The fundamental difference lies in the representation used: do we perform the subtraction in the wavelet space or in the standard $y - \phi$ representation? So, although the underlying philosophy is, to a large extent, the same for all methods, the value of the wavelet representation may be assessed by comparing wavelet and standard methods, which use the same tools[13]. Such a detailed comparison is not performed in this study, where we instead focus on what can be achieved by wavelet-based method in their own right.

[12] Although something along these lines might, to some extent, be possible

[13] For instance, wavelet-based subtraction which uses no auxiliary information, compared to PuPPI; or wavelet-based subtractions which use track information, compared to track-guided PuPPI or NpC; etc.

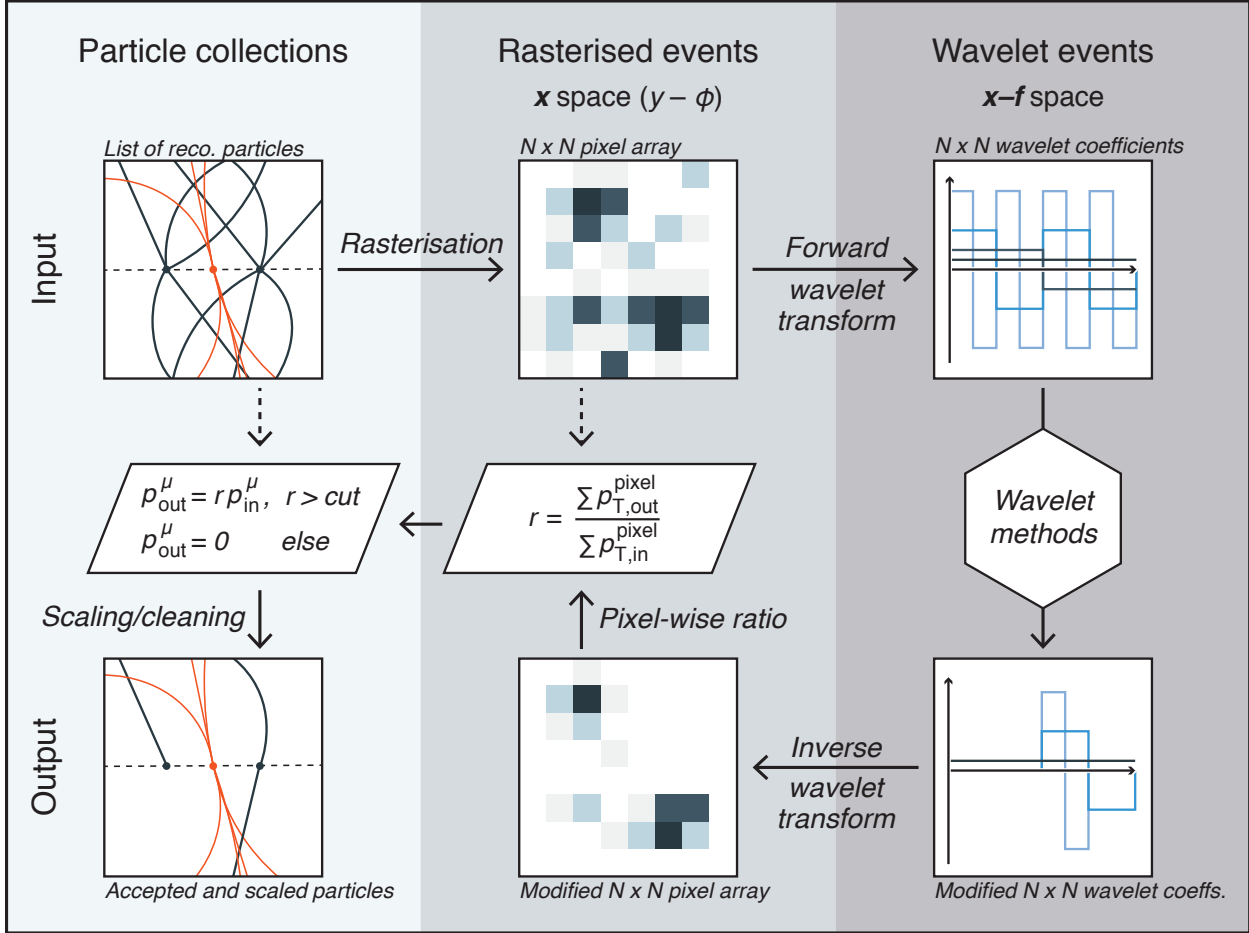| Particle collections	Rasterised events	Wavelet events

Figure 2.6: Overview of wavelet analysis. A list of input physics objects is rasterised into an $N \times N$ pixel array, which is forward transformed into the $N \times N$ space of wavelet coefficients (wavelet event). After modification by some wavelet methods the event is inversely transformed and the resulting rasterised event is compared to the input one. Based on the per-pixel output-to-input ratio of summed $p_\perp$, the input list of particles is reduced (if the ratio is below some threshold) and scaled (if the ratio is above this threshold). Position (x) space refers to $y - \phi$ space, and position-momentum ($x - f$) space refers to the space of wavelet indices ($m_{y,\phi}, l_{y,\phi}$).

[14] This naming convention is consistent with the one used in NEWWAVE [67].

■ Applying wavelets

We have now given a short motivation for why wavelets might prove useful under conditions with large amounts of soft QCD activity, and now we will turn to the practical matter of applying wavelet-based analyses to hadron collision events. In the interest of consistency, we propose a short glossary, defining terms which will be used frequently in the following, which can hopefully serve as a starting point for a coherent discussion about wavelets in high energy physics[14]. Incidentally, the structure of the glossary also follows the steps taken in actually applying wavelet methods to collision events, which is depicted in Fig. 2.6.

Dictionary. The set of wavelet and scaling functions used, forming the basis of the wavelet transform. The dictionary might be chosen from pre-existing/-implemented ones. For instance, the Haar, Daubechies $D_{N \leq 20}$, and biorthogonal spline wavelet families are implemented in GSL [65]. The 2D Haar basis function were shown already in Fig. 2.3 for an 8×8 pixel array. The choice of dictionary might very well depend on the specific needs of the analysis, as well as considerations regarding complexity, boundary conditions, etc. A common metric for the usefulness of a particular dictionary is the *sparsity* of the data in this representation, as measured by the

steepness of the energy spectrum of the coefficients for the basis: a dictionary where very few coefficients carry almost all of the energy/information about the data is usually considered desirable (particularly in regards to compression of data). Curiously, the sparsity of the standard pixelated $y - \phi$ representation is given by the pixel $p_\perp$ distribution. However, size data size is not our main constraint in the present study, we will not use sparsity as decisive factor.

Alternatively, it might be possible to construct custom wavelet families, specifically suited for representing high energy hadron collisions, e.g. building on the structure of the splitting kernels in Eq. (1.40). However, this line of inquiry has currently not been pursued further.

Pixel definition. The grid in $y - \phi$ space defining the 2D event display, or "image", to which we apply the wavelet transform. One should specify the extent of the pixel definition in $|y|$ and ϕ, the latter always being $[0, 2\pi[$, as well as the resolution $N \times N$ of the pixel array, under the condition that side lengths should be an integer power of 2, i.e. $N = 2^n$ for some integer n. This requirement is a direct consequence of the dyadic scaling nature of the wavelet transform, cf. Sec. 2.1.

Input collection. The physical objects which one wants to study/modify using wavelet methods. Depending on purpose, different reconstructed objects may be used: calorimeter clusters, particle flow objects, charged particle tracks, etc.

Rasterisation. The procedure of iterating all input objects and adding the scalar $p_\perp$ of each object, covered by the pixel definition, to the pixel within which it falls.

Rasterised event. The result of the rasterisation: a 2D event display, spanned by the pixel definition, the height of each pixel being equal to the scalar sum of $p_\perp$ deposited within its $y - \phi$ extent.

(Forward) wavelet transform. Done using the specified dictionary, by performing forward 2D discrete wavelet transforms (DWT), i.e. a forward 1D DWT on each row in the rasterised event, followed by an forward 1D DWT on each ensuing column. See Sec. 2.1 for details.

Wavelet event. The $N \times N$ space of wavelet coefficients resulting from the (forward) wavelet transform. This is the space upon which we perform our *wavelet methods*, modifying the individual wavelet coefficients. This space is akin to the momentum space of the Fourier transform, the fundamental difference being that the indices of the wavelet basis functions encode frequency *as well as* position information. This can be seen in Fig. 2.3.

Wavelet coefficient. Each of the entries in the wavelet event, specifying the contribution of each wavelet basis function. Each coefficient

$C_{m_{y,\phi};l_{y,\phi}}$ has a value with units of momentum and unique indices specifying its scale ($m_{y,\phi}$) and translation ($l_{y,\phi}$) in both y- and ϕ-direction. See Sec. 2.1.

Wavelet method. Any operation applied to the wavelet event, e.g. to mitigate the impact of pile-up. A few possible methods are proposed in Sec. 2.3.

Inverse wavelet transform. Done by performing an inverse 2D DWT, i.e. an inverse 1D DWT on each column in the wavelet event, followed by an inverse 1D DWT on each ensuing row. This results in a (modified) rasterised event. As noted before, setting all but one wavelet coefficient to zero and performing the inverse transform yields the corresponding wavelet basis function in $y - \phi$ space. These are shown in Fig. 2.3 for the Haar wavelet.

Output collection. The (modified) rasterised event resulting from the inverse wavelet transform is compared to the original rasterised event: for each pixel, if the ratio of its value to that of the corresponding pixel in the original rasterised event is larger than some threshold grater than or equal to[15] 0, scale[16] the momentum of all objects in the input collection falling inside this pixel by the this ratio; otherwise all objects within this pixel are discarded. The resulting reduced and modified collection of physics objects is called the output collection.

Wavelet analysis. The combined process depicted in Fig. 2.6, of reducing and re-scaling a list of input physics objects based on a (set of) wavelet method(s).

The wavelet analysis is performed on an event-by-event basis, where the input object collection changes for each event, but the underlying pixel definition is the same for all events. Similarly, the methods applied to the wavelet event might be exactly the same for each event, or they might change depending on some event-specific information, e.g. the collection of tracks or the measured number of pile-up vertices.

■ Particle scaling

Following the steps in Fig. 2.6 up to the modified rasterised event, we have performed an almost purely mathematical exercise. However, when taking the last step—of re-scaling the physical particles—we are modifying physical observables, hopefully in a meaningful way .

Mass scaling

This modification involves scaling the four-momenta of the physics objects as

$$p^{\mu}_{\text{out}} = r p^{\mu}_{\text{in}} \tag{2.43}$$

[15] Ratios < 0, as well as ratios > 1 may occur, depending on how the wavelet event is modified.

[16] By scalar multiplication of the ratio with the object's momentum four-vector, see Fig. 2.6.

which has the implication that

$$m_{\text{out}} = \sqrt{p_{\text{out}}{}^\mu p_{\text{out}\mu}} = \sqrt{r^2 p_{\text{in}}{}^\mu p_{\text{in}\mu}} = r m_{\text{in}} \approx r m_{\text{phys}} \qquad (2.44)$$

which, using physics objects with fixed, non-zero pole masses (e.g. re-constructed electrons) is problematic. However, if we're only re-scaling calorimeter-level inputs, like topological clusters which are massless by definition, this problem is avoided. Even if one *were* to modify massive objects, the methods proposed in the following would not be invalidated: one just has to make clear, that the output collection doesn't contain physical particles which are expected to be on shell, but rather four-vector *proxies* which, on the whole, are expected to represent the hard scatter event as it would have been measured in terms of actual, physical particles.

Particle cleaning (CL)

An interesting point is, that when re-scaling the input objects, a minimal, physically necessary requirement is, that the scaling factor should be ≥ 0. Otherwise we would obtain particles with negative energies and masses, cf. Eq. (2.44), which would be unacceptable. However, we are free to impose a threshold on the pixel-wise ratio which is greater than zero and thus discard more particles than with the minimal threshold. This can be motivated as follows:

The wavelet methods that we apply to the wavelet event will, in some way or another, be designed specifically to remove some part of the event. In the present study, we are concerned with mitigating the effects of pile-up, and so we will use methods designed to remove the (expected) contribution from pile-up to the event as a whole. When we then perform the inverse transforms, some pixels will have had their values changed a lot, while others will have hardly changed at all. Since the methods are intended to remove the contributions from pile-up in the space of wavelet coefficients, their effect in the rasterised event is expected to be the same. Since the question of whether certain particles come from pile-up is binary[17], it is meaningful to discard reconstructed objects if we have sufficient reason to believe that they are indeed from pile-up. This is the principle behind SoftKiller, cf. Sec. 1.4. Therefore, we may decide that pixels with an output-to-input ratio below some non-zero threshold exhibit a behaviour which we would expect from pile-up, and thus discard all objects within this pixel. Pixels with an ouput-to-input ratio above this threshold will not have changed sufficiently to be labeled as being attributable to pile-up, and so all particles within these are kept, with their four-momenta scaling in the usual way, cf. Eq. (2.43). The hope is then, ideally, that the resulting list of (scaled) particles closely resembles the "true" set of particles coming from the hard scatter interaction, including UE.

Regardless of the choice of wavelet methods, we can always choose a subsequent non-zero re-scaling threshold. In the following, we will label this process *particle cleaning* (CL), and think of it as a "meta-method", which may be used in combination with other methods.

[17] That is, non-probabilistic: either their do or they do not, with no interference with the hard scatter interaction.

2.3 Wavelet methods

At this point, we have outlined the theory for the (discrete) wavelet transform, motivated their use hadron collisions, and described the steps taken in applying them to individual collision events, but we still have not provided any examples of wavelet methods, despite their being front and center in the whole concept of a wavelet analysis. This is what we turn to next. In the following, we will describe and motivate various wavelet techniques, some of which are used in the following study. We will also suggest a general wavelet analysis structure, intended to guide the way and order in which we apply individual wavelet methods.

■ Proposed methods

Below, we detail a small selection of wavelet methods, intended to mitigate the effect of pile-up and underlying event (UE) on hard scatter events. Please note, however, that these are just a first collection in a (hopefully) ever expanding toolbox of techniques, intended both for general and specialised use. All methods are presented as they appear in Fig. 2.6, taking a wavelet event as input and returning a wavelet event with modified coefficients.

Flat de-noising (FDN)

In any efficient (sparse) representation, we expect a small number of coefficients to carry most of the information content (i.e. have large energies) and therefore we only need a small number of coefficients to represent the significant characteristics of the data with a minimum of loss. This is the concept behind image compression. Therefore, all of the coefficients which have comparatively small energies or information contents (here: absolute values) are expected to have a negligible contribution to the representation of the whole event, this contribution being in the form of "soft noise". Therefore, one method to remove such soft noise (e.g. from the UE) would be to reject all coefficient with an absolute value below some fixed threshold, or cut, by setting them to zero. This method is called *flat de-noising* and is possibly the simplest method one can apply, as it is a standard technique known e.g. from image de-noising. This method is implemented as shown in Fig. 2.7.

Figure 2.7: Pseudo-code implementation of the flat de-noising method.

```
1  given waveletEvent, cut
2  for each coefficient in waveletEvent do
3      if | coefficient .value| < cut then
4          coefficient .value = 0
5      end
6  end
```

   The remaining high-energy coefficients are expected to constitute a cleaner representation of the hard part of the event, which is what we seek. The value of the threshold used is not fixed *a priori*, but

is expected [67] to be on the order of $\mathcal{O}(\Lambda_{\mathrm{QCD}}) \approx 0.1$–$1\,\mathrm{GeV}$, since then QCD becomes non-perturbative and the parton showering model gives way for hadronisation. Therefore, this is the scale at which the soft QCD from either UE or pile-up is expected to dominate.

Scale-dependent de-noising (SDN)

The flat de-noising method makes no assumptions as to the behaviour of pile-up in the space of wavelet coefficients. Therefore it might be expected to have sub-optimal performance in regards to our goal of removing the (expected) contribution from pile-up to the total event. A slightly more complex methods is the *scale-dependent de-noising*, which operates similarly to the flat de-noise methods, but with coefficient thresholds depending on the scale indices of the coefficients (i.e. on the size in $y - \phi$ of the corresponding wavelet basis function). The cuts are given by [67]

$$\mathrm{cut}_{m_y, m_\phi} = N_{\mathrm{PU}} \times \frac{2\,y_{\mathrm{max}}}{2^{m_y} \times 2^{m_\phi}}\,\mathrm{GeV} \qquad (2.45)$$

These cuts are motivated [67] by the fact that each mininum bias collision contributes approx. $1\,\mathrm{GeV}$ per unit rapidity, cf. Sec. 1.4, resulting in an expected energy from pile-up of $\approx N_{\mathrm{PU}} \times 2\,y_{\mathrm{max}}\,\mathrm{GeV}$ in an event with N_{PU} pile-up collisions and a rapidity range of $|y| < y_{\mathrm{max}}$. This value is then scaled by the approximate fractional area, $1/(2^{m_y} \times 2^{m_\phi})$, that the wavelet basis functions at scales (m_y, m_ϕ) take up, see e.g. Fig. 2.3. This method is implemented as shown in Fig. 2.8.

```
1  given waveletEvent, N_PU
2  for each coefficient in waveletEvent do
3      m_y   = coefficient.m_y
4      m_phi = coefficient.m_phi
5      cut  = N_PU × 2 y_max / 2^(m_y + m_phi) GeV
6      if | coefficient .value| < cut then
7          coefficient .value = 0
8      end
9  end
```

Figure 2.8: Pseudo-code implementation of the scale-dependent de-noising method.

Whereas the flat de-noise method is intended to remove diffuse noise, regardless of origin, the scale-dependent de-noise method is intended to specifically remove the contribution from pile-up, by using cuts which are meant to follow the actual distributions of wavelet coefficients in pile-up, based on purely geometrical arguments. We note, that the proposed implementation is independent of the translation indices $l_{y,\phi}$ of the coefficients, which means that all wavelet coefficients at the same set of scales are subjected to the same cut, regardless of their physical position in the $y - \phi$ space of the detector.

Track filtering (TF)

So far, both methods presented have been purely geometrical, or at least used no auxiliary information to guide the cuts applied to the full

wavelet event, apart from N_{PU}. Both of these methods also apply cuts which are identical for all coefficients at each set of scales (m_y, m_ϕ). However, we might reasonably expect that the optimal choice of which coefficients to keep and which to discard to be dependent on whether a particular basis function is located in a high-energy density region of the detector (e.g. in the vicinity of a jet) or in an otherwise empty region. Put in another way: having made the case for the wavelet function being localised, we might as well attempt to use this feature creatively.

The *track filtering* method is intended as a dedicated pile-up removal tool, similarly to the scale-dependent de-noising. It utilises the fact, that one can—with little ambiguity—associate inner detector tracks to the vertex from which the they originated. This means that we for each collision, up to tracking resolution, efficiency, etc., can construct two additional wavelet events, apart from the wavelet representation of the full event ('FE'): one from tracks from the hard scatter only ('HST'; i.e. those associated to the primary vertex), and one from all tracks in the full event, regardless of their origin ('FET').

The assumption is then that, in the wavelet domain, tracks alone faithfully encode information about the combined event, of which they are a part. That is, that the wavelet representation of tracks coming from the hard scatter interaction also describe the neutral particles from this interaction; similarly for pile-up interaction We stress that this is an *assumption*; one whose validity we must later test, and one which has previously been put forward in the $y - \phi$ representation e.g. by the NpC method, cf. Sec. 1.4, where it did not prove as successful as one might have hoped. Seeing as our ultimate goal is to estimate the coefficients of the hard scatter event as a whole ('HS'), including neutral particles, as it would have occurred if it could be measure in the absence of pile-up, this assumption can be cast in the form of Eq. (2.46)

$$\left| \frac{C^{\mathrm{HS}}_{m_{y,\phi};l_{y,\phi}}}{C^{\mathrm{FE}}_{m_{y,\phi};l_{y,\phi}}} \right| \approx \left| \frac{C^{\mathrm{HST}}_{m_{y,\phi};l_{y,\phi}}}{C^{\mathrm{FET}}_{m_{y,\phi};l_{y,\phi}}} \right| \tag{2.46}$$

That is, for each basis function, we expect the ratio of the corresponding coefficients in the hard scatter event to the full event to behave similarly to the ratio of the same-indiced coefficients in the wavelet event of hard scatter tracks to that in the wavelet event of all tracks. This is by no means the only way in which the assumption of tracks-to-neutrals correlations can be formulated, but it is one which resembles Eq. (1.61) from the NpC method and one is easily implemented as a wavelet method. This is done as follows.

If the ratio $|C^{\mathrm{HST}}_{m_{y,\phi};l_{y,\phi}} / C^{\mathrm{FET}}_{m_{y,\phi};l_{y,\phi}}| \approx 1$, the hard scatter contribution to this particular coefficient in the full event, $C^{\mathrm{FE}}_{m_{y,\phi};l_{y,\phi}}$, is dominant, and we keep the coefficient. Conversely, if $|C^{\mathrm{HST}}_{m_{y,\phi};l_{y,\phi}} / C^{\mathrm{FET}}_{m_{y,\phi};l_{y,\phi}}| \approx 0$, the coefficients value is dominated by energy *not* associated to the hard scatter, i.e. from pile-up, and the coefficient is discarded. In general, we introduce some (fixed) cut, such that all coefficients, for which

the hard scatter-to-full event track ratio falls below this threshold, the corresponding coefficient in the full event is discarded; otherwise it is kept unmodified.

The only exception we make is for the case where $(m_y, m_\phi) = (0, 0)$. As seen in Fig. 2.3, this basis function measures the total average energy deposited in the event. For this particular set of scales, the argument that the ratio of the average energy carried by hard scatter tracks to that of all the tracks in the full event should be close to unity doesn't hold, especially in situations with a large number of pile-up collisions. Rather, we expect it to be $|C_{0,0;0,0}^{\mathrm{HST}}/C_{0,0;0,0}^{\mathrm{FET}}| \approx E_{\mathrm{tot.}}^{\mathrm{HS}}/E_{\mathrm{tot.}}^{\mathrm{FE}}$, cf. Eq. (1.61), which is strictly smaller than one. Therefore, completely discarding this basis function seems unjustified[18], and it is thus always left unmodified by the track filtering method. It might be possible, and justified, to *scale* $C_{0,0;0,0}^{\mathrm{FE}}$ by the ratio $|C_{0,0;0,0}^{\mathrm{HST}}/C_{0,0;0,0}^{\mathrm{FET}}|$; however, this is not done for the present method, but see below. This method is implemented as shown in Fig. 2.9.

[18] What is more, removing the average energy in the event has critical impact on e.g. jet energy measurements, and should therefore be done only will great care.

```
1   given waveletEvent_FE, waveletEvent_HST, waveletEvent_FET, cut
2   for each coefficient_FE in waveletEvent_FE do
3       if coefficient_FE.m_y == 0 and coefficient_FE.m_phi == 0 then
4           do nothing
5       else
6           coefficient_HST = get same coefficient as coefficient_FE in waveletEvent_HST
7           coefficient_FET = get same coefficient as coefficient_FE in waveletEvent_FET
8           if coefficient_FET.value == 0 then
9               do nothing
10          else if | coefficient_HST.value / coefficient_FET.value| ≤ cut then
11              coefficient_FE.value = 0
12          end
13      end
14  end
```

Figure 2.9: Pseudo-code implementation of the track filtering method.

We see that two special situations may occur: if the coefficient in wavelet event of all tracks is equal to zero, it means that for this set of indices we have no track information to guide our choice; therefore we choose to be conservative and leave the coefficient unaltered. In this case it might be possible to use a generic, geometrically motivated fall-back cut; however, this has not been investigated further in the present study. Similarly, we use a smaller-than-or-equal-to criteria when imposing the cut, rather than a strict inequality. This means that, even in the case where cut $= 0$, if a particular coefficient has no contribution from tracks associated to the hard scatter, but does have contributions from other, non-PV tracks, the coefficient is thought to have no neutral hard scatter contribution either, and is killed. Therefore, choosing a cut of zero has a non-trivial effect on the output event. However, one may choose to use a strict inequality in the implementation, whereby TF(0) becomes a trivial identity operator, similar to FDN(0) (modulo particles falling outside the pixel definition being discarded, regardless of the choice of method).

Note, that this method chooses which coefficients to keep in a way which differs markedly from the previous methods: this method uses information from tracks to *guide* which wavelet coefficients are kept and which are discarded. Therefore, the cut is not imposed on the value of the wavelet coefficient in question (which is not considered *at all*), as opposed to the two previous methods, but rather on the ratio of the same-indiced coefficients in the wavelet events for tracks alone.

Track scaling (TS)

The *track scaling* method is based on exactly the same logic as track filtering, but takes the assumption in Eq. (2.46) one step further. In stead of using the hards scatter-to-full event track ratio to just *guide* which coefficients in the full event to discard, we now cast it in the following form

$$C^{\text{HS}}_{m_{y,\phi};l_{y,\phi}} = C^{\text{FE}}_{m_{y,\phi};l_{y,\phi}} \times \left| \frac{C^{\text{HST}}_{m_{y,\phi};l_{y,\phi}}}{C^{\text{FET}}_{m_{y,\phi};l_{y,\phi}}} \right| \tag{2.47}$$

and instead *force* the coefficients of the full event to follow the structure of the track events. We note that, contrary to the case with track filtering, for track scaling the case with $(m_y, m_\phi) = (0,0)$ is not a problem—actually quite the opposite. Here, the average energy in the full event is scaled by the ratio of the average energy of the hard scatter tracks to that of the full event tracks, which is perfectly reasonable: if the ratio of charged to neutral particles is reasonably constant (~ 2) and their $p_\perp$ spectra are reasonably similar, we *do* expect the average energy of the hard scatter tracks to be representative of the energy of the hard scatter event as a whole. Therefore, scaling the coefficient $C^{\text{FE}}_{0,0;0,0}$ by the ratio $|C^{\text{HST}}_{0,0;0,0}/C^{\text{FET}}_{0,0;0,0}| \approx E^{\text{HS}}_{\text{tot.}}/E^{\text{FE}}_{\text{tot.}}$ makes perfect sense, and this exception is dropped. The method is implemented as shown in Fig. 2.10.

Figure 2.10: Pseudo-code implementation of the track scaling method.

```
1   given waveletEvent_FE, waveletEvent_HST, waveletEvent_FET, cut
2   for each coefficient_FE in waveletEvent_FE do
3       coefficient_HST = get same coefficient as coefficient_FE in waveletEvent_HST
4       coefficient_FET = get same coefficient as coefficient_FE in waveletEvent_FET
5       if  coefficient_FET.value == 0 then
6           do nothing
7       else then
8           ratio  = | coefficient_HST.value / coefficient_FET.value|
9           coefficient_FE.value = coefficient_FE.value × ratio
10      end
11  end
```

This implementation strains the assumption of tracks-to-neutrals correlation in wavelet space to its maximum, and e.g. entails scaling individual coefficients by ratios which may be greater than one. Forcing the full event wavelet coefficients in this way is quite drastic and should be done with care and thorough justification.

■ Choice of passes

In the preceding sections we have consistently talked about wavelet methods in plural, since each method described above, and any other method one might see fit, can be performed in succession, as passes or stages in a full wavelet analysis, quickly yielding a daunting number of possible combinations. To guide how we apply the methods, we propose a general three-stage structure[19]:

[19] Suggested by J. Monk.

(1) Pile-up mitigation. We impose cuts on, or scale, each wavelet coefficient depending on the expected contribution from pile-up to this particular coefficient. This expectation can be based on either geometrical/average considerations (e.g. for 'scale-dependent denoise') or on auxilliary information such as tracks (e.g. for 'track filtering').

(2) Clean-up. Following the remarks made in Sec. 2.2, when we perform the inverse transform, the individual particles are scaled by an amount which is intended to reflect the contribution from pile-up to the particular pixel within which they fall. As was mentioned, since the methods in pass (1) are intended to remove pile-up, we apply the 'particle cleaning' meta-method, and impose a non-zero threshold when performing the re-scaling of the physics objects in the input collection. This ideally leaves us with a collection of (scaled) objects coming from the hard scatter process alone.

(3) Underlying event mitigation. We can then perform another wavelet transform, using the skimmed and re-scaled particle collection as input, and impose additional cuts on the (new) wavelet coefficients depending on the expected contribution from the UE. Since the question of whether an objects originates from the UE is probabilistic in nature[20], we perform the inverse transform, and scale the momentum of each object corresponding to the amount of soft UE activity it is thought to carry, but we do not discard individual particles altogether beyond the physically minimal re-scaling threshold of zero. This is justified by the assumption that all objects that pass (1) and (2) originate from the hard scatter event.

[20] There is no clear, meaningful way, at a quantum mechanical level, to distinguish between what is "hard" and what is "soft" in any given process, in contrast to the case of pile-up.

III   Analysis: Toy simulation

3.1 Physics case and setup

In the previous chapter we introduced the theory of wavelets, suggested their use in a hadron collision context, and described and motivated a few wavelet methods intended to mitigate the effect of pile-up. We now turn to the questions of which which metrics to use in assessing the effectiveness of the methods. We then discuss which physics cases most suitable for studying these metrics, and how we generate samples of them in practice. We describe the smearing procedure using in the following sections, intended to crudely mimic the performance of the ATLAS detector. Finally, we describe the wavelet setup and jet clustering scheme we will be using in the following sections.

■   Metrics

So far, we have presented a few wavelet methods, but we still need to decide on a metric on the basis of which to asses their effectiveness. As we have explained, the main goal of the present study is to assess whether, and to what degree, wavelet-based methods can be employed to improve sensitivity in searches for new physics; in particular, by mitigating the effects of pile-up in particular and soft QCD activity in general. We noted, that this work is thought to be particularly valuable to physics analyses relying on jets, since their energy and mass resolution are crucial e.g. for the sensitivity[1] in high-mass resonance search beyond the Standard Model [14, 15, 16]. In the following, we will be using two complementary metrics, in order to study the impact of wavelets on different aspects of jets reconstruction.

[1] For instance, the natural width of the W and Z bosons are $\Gamma_W = 2.085\,\text{GeV}$ and $\Gamma_Z = 2.4952\,\text{GeV}$ [2], while the width of the experimentally reconstructed W/Z boson jet mass distributions are $\sigma_{M_{\text{jet}}} \approx 8 - 9\,\text{GeV}$ [52].

Jet energy resolution. In the general case of reconstructing jets, regardless of their origin, we will be interested in measuring the energy (or $p_\perp$) of the jet as reliably as possible, assuming it to be a good measure for the energy of the initiating particle. A good jet energy resolution is important in searches using jets as well as in the measurement of missing transverse energy [68]. Furthermore, since the $p_\perp$ spectrum of QCD jets falls rapidly as $\sim p_\perp^{-a}$ with $a \approx 6$, cf. Sec. 1.2 and Fig. 3.1, a relative error on $p_\perp$ of $\delta p_\perp / p_\perp$ propagates to a relative error on the jet count of $|a|\delta p_\perp / p_\perp$, i.e. increasing by a factor of roughly 6.

This large error has the effect of limiting sensitivity in jet-based search, cf. e.g. [14] where the jet $p_\perp$ *resolution* is the all-dominating systematic uncertainty affecting the shape of the reconstructed signal (here: diboson mass width). Similarly, any *bias* in the measurement of the jet energy, brought about by the presence of pile-up, has the effect of biasing its reconstructed kinematic variables. Therefore we will be using jet energy resolution and jet energy bias as one metric for the usefulness of our methods. Specifically, this is done by applying our wavelet methods to the same hard events, with and without overlaid pile-up; matching the reconstructed jets, with and without pile-up, in $y - \phi$; and studying the mean and standard deviation of the relative difference in energy of matched
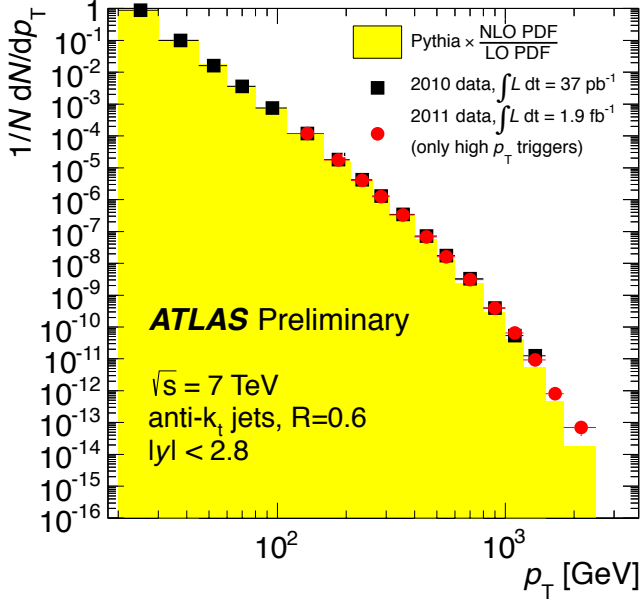
jets, which is identified as the contribution to the energy which can be attributed to pile-up,

$$\frac{E_{\mathrm{jet}}^{\mathrm{w/\,PU}} - E_{\mathrm{jet}}^{\mathrm{w/o\,PU}}}{E_{\mathrm{jet}}^{\mathrm{w/o\,PU}}} \tag{3.1}$$

i.e. the part which we want to remove.

Note, that here we are not considering detector effects like overall jet energy scale and the non-compensating nature of the ATLAS calorimeter system. We are only concerned with mitigating the effects from pile-up.

Boson jet mass. When looking at (boosted) boson jets, the main characteristic separating them from their non-resonant counterparts (henceforth called 'QCD jets' for short) is their invariant mass, M_{jet}, which is hoped to reflect the pole mass of the boson in question. This fact is used extensively in searches involving boson jets in the final state [14, 15, 16], where the jet mass cut—usually symmetric around the boson pole mass(es)—is vital in killing the otherwise overwhelming QCD background. Therefore, being able to reconstruct the mass of boson-initiated jets precisely and in an unbiased manner has the potential to significantly enhance the sensitivity in such searches. In order to take into account both the narrowing of the boson jet mass peak, the possible loss of efficiency brought about by the wavelet methods, and the possible removal of QCD jets around the W/Z pole mass, we will be using $S/\sqrt{S+B}$ as our metric, or figure of merit. Here, S is the number of boson jets within some M_{jet} range and B is the number of QCD jets with the same range. As we will usually be concerned with situations in which $B \gg S$, we will in the following use the simplified form $S/\sqrt{B}$.
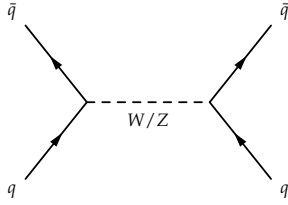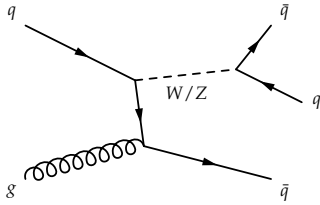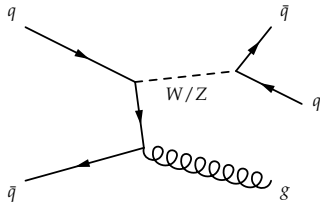
■ Processes studied

Having settled on the metrics which will guide our work, we must decide on which physics processes permit the assessment of these. Firstly, we will need samples containing (boosted) hadronically decaying vector bosons, since we are motivated by the decay of exotic resonances to diboson final states. We choose to use Standard Model single boson continuum production (henceforth referred to as 'V + jets'), since this process is the most commonplace (relative to continuum diboson production) and well-understood (relative to e.g. $H \rightarrow WW/ZZ$). Leading order (LO) $2 \rightarrow 1$ W/Z-production is achieved through direct Drell-Yan production, shown in Fig. 3.2.

However, the $p_\perp$-scale of the boson will, to lowest order, be given by the "primordial" $k_\perp$ of the initial-state quarks [69]. Parton showers will generate additional jets up to moderate $p_\perp$, but since we are interested in the very high-$p_\perp$ tails of the distributions we need to include next-to-leading-order (NLO) effects from at least one additional parton leg, by means of $2 \rightarrow 2$ matrix elements [23]. The NLO production diagrams are shown in Fig. 3.3.

An additional benefit of using continuum single-boson production is, that requiring the boson jet to have a large $p_\perp$ implies, by conservation of momentum in the transverse plane, that it will be accompanied by a equally high-$p_\perp$ jet (henceforth referred to as the 'recoil jet'), initiated by the hard parton radiated in the initial state. The presence of jets initiated by a hardonically decaying, resonant vector bosons *as well as* a jets initiated by non-resonant partons in the same event will in allow us to simultaneously study the impact of wavelet methods on jets of different origin.

However, in order to study the effect of wavelets on QCD jets, with a composition of initiating partons which is true to what would be observed in data, we have chosen to also generate dedicated samples of continuum QCD dijets. The main, topologically distinct production diagrams, drawn in the s-channel, are shown in Fig. 3.4.

Notice here, that the leading order diagrams suffice to obtain high-$p_\perp$ jets, since each final state parton will hadronise into a boosted jet with $p_\perp$ given by roughly half the invariant mass of the di-parton system. However, as in the case for V + jets production, parton showering will serve to simulate the effect of additional parton legs.

■ Generated samples

Having outlined which physics processes we will be using, we must decide on how to generate the events. Since PYTHIA8 [70] includes $W/Z + 1$ parton matrix elements, in addition to parton showering, we will use this generator to simulate our hard processes, both V + jets and QCD dijets. For the hard events, we have used PYTHIA8.205 with the A14-NNPDF23LO tune [71], which is tuned to the full set of 7 TeV data collected by the ATLAS experiment ("ATLAS 2014") in order to perform well for high-$p_\perp$ Z bosons and dijet events. The tune used here is based on the NNPDF 2.3 LO PDF set, which is an upgraded

Figure 3.2: Leading-order (Drell-Yann) W/Z production diagram.

(a) qg production.

(b) qq production.

Figure 3.3: Next-to-leading-order production diagrams for $W/Z(\rightarrow q\bar{q})$ + jets at pp colliders. Additional u-channel diagrams are implied.

version of the NNPDF 2.1 LO set used in Fig. 1.8 [72]. In order to have enough statistics at high $p_\perp$, we generate three sets of samples with generator-level cuts on the boson $p_\perp$ of 180, 280, and 600 GeV. Due to the effects of pile-up and UE, this translates to reconstruction-level $p_\perp$ cuts of 300, 500, and 1000 GeV, resp. These slices cover jet $p_\perp$ up to ~ 2000 GeV, which is also the region of interest for current analysis in diboson final states [14, 15, 16], see also Sec. 1.1.

These $p_\perp$ cuts apply to (groomed) jets clustered in the unmodified event. After applying a wavelet analysis of our choice, we perform a similar clustering with a much lower (50 GeV) $p_\perp$ cut, and match the jets to those in the unmodified event within $dR < 0.6$.[2] Jets clustered in the wavelet-subtracted event, which are matched to a reconstructed jet in the unmodified event, are kept for further study; the rest are discarded. We choose this jet selection, since the reconstruction-level $p_\perp$ cuts are chosen to take into account the energy contribution from pile-up; but considering that wavelet analyses will aim to remove exactly this component, imposing the same $p_\perp$ cut on jets clustered in the wavelet-cleaned event would introduce a misleading inefficiency. Therefore, any $p_\perp$ requirement referenced in the following applies to jets clustered in the unmodified, where it is understood that jets in the wavelet-cleaned event are matched to these in dR.

The hard events are overlaid with soft QCD events (minimum bias, or minbias) using the PILEMC package [73]. The nominal set of pile-up events used are also generated with PYTHIA8.205, however, using the A2-MSTW tune [74], which is a minbias tune based on the MSTW 2008 LO PDF set. The number of minbias events generated correspond to ca. 50 per hard events, and they are overlaid in order, according to a Poisson distribution around some average number for pile-up events, $\langle \mu \rangle$. Since we will be studying pile-up multiplicities up to $\langle \mu \rangle = 200$, we will be re-using some of the same pile-up events for different hard events. This is mainly a matter of convenience, and hopefully the effect on the final results should be negligible. This is supported by the fact that we are studying jets, which are localised in relatively small regions of the full detector area. This means that, even though we use the same minbias event multiple times (and in the same order), we are going to sample them in different regions of $y - \phi$ space, depending on the direction of the bosons/final state partons in the hard event.

If one were to take this study further, one might us a larger collection of minbias events, sample them randomly instead of in order, and perhaps use additional high-$p_\perp$ slices in order to have larger statistics for the cases which are expected to influence the results the most.

All of the analysis in the following sections has been performed in the RIVET framework [75], interfaced with ROOT [76], to allow for easy transferability. The wavelet analysis is implemented as a custom Rivet::FinalState, configurable to use any (combination) of the wavelet methods proposed in Sec. 2.3. These are implemented using the NEWWAVE package [67] which in turn relies on the GNU

(a) $gg \rightarrow gg$

(b) $qq \rightarrow qq$, quartic

(c) $qg \rightarrow qg$

(d) $qq \rightarrow gg/gg \rightarrow qq$
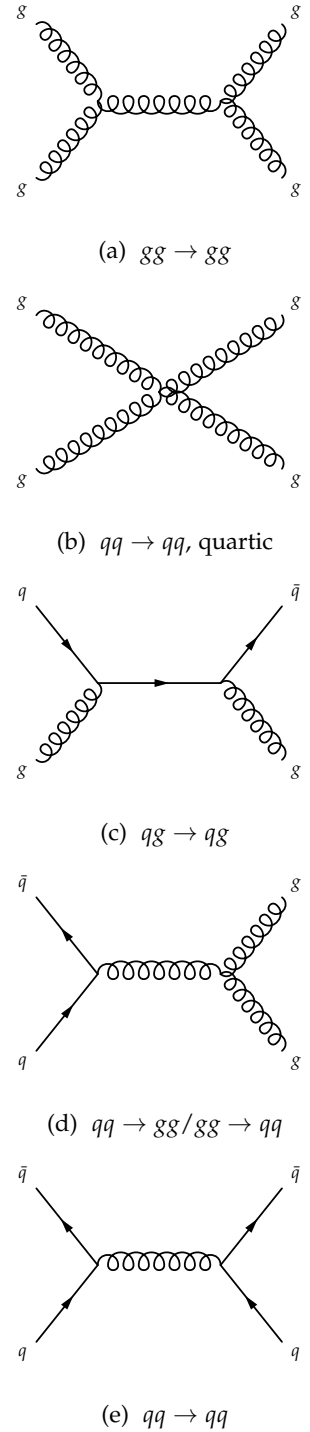

(e) $qq \rightarrow qq$

Figure 3.4: Leading-order, topologically distinct, production diagrams $2 \rightarrow 2$ QCD processes, used in the simulation of inclusive QCD dijet samples.

[2] This value was found in to be suitable by studying distributions of dR_{jj}, between modified and un-modified events, in simulation.

Scientific Library (GSL) [65] to perform the actual wavelet transforms.

Since this study in conducted in a time where the LHC is going into Run 2the, all samples used in the following with be generated at a center of mass energy of $\sqrt{s} = 13\,\text{TeV}$, since all 7 and 8 TeV analyses will have concluded or are nearing conclusion at this point. Similarly, as remarked in Sec. 2.2, we will mainly be interested in run conditions with a large amount of pile-up, and consequently our nominal setup will be using $\langle \mu \rangle = 40$, but we will be studying average pile-up multiplicities up to 200.

■ Smearing

In order to make at least a crude comparison with official Monte Carlo simulation and, in turn, actual data, we need to smear the generator-level particles in a way so as to resemble the effects of a physical detector. Since we will be comparing with official ATLAS simulation, we impose the $|\eta|$ coverage and smear the particles by the relevant nominal resolutions as listed in Tab. 1.4 [32].

All particles have their momentum changed by a gaussian smearing centered at zero and with a width given by the appropriate resolution in Tab. 1.4: charged particles are identified as tracks and smeared by the ID resolution; neutral *and* charged particles (except muons) are identified as calorimeter deposits, and for convenience these are all smeared by the HCAL resolution[3]. This means that the ECAL resolution is not used, which also constitutes a worst-case-scenario for the smearing. If the smearing results in a $p_\perp$ or E value which is negative, the corresponding variable is set to zero. Since muons—and in particular high-$p_\perp$ ones—will not occur in our chosen samples, since they are not central to the present jets-only study, these are smeared like tracks, i.e. disregarding the effect of MS measurements. Finally, neutrinos escapes the detector undetected.

Since the ATLAS detector has a lower limit on the energy of objects which can be reconstructed, we impose a lower tracking $p_\perp$ threshold of 400 MeV, just below the nominal threshold of 500 MeV [32] and comparable to that of the other LHC experiments. However, the ATLAS tracking system has a track reconstruction efficiency below unity even for tracks passing the lower $p_\perp$ threshold, as shown in Fig. 3.5. This mainly due to the material in the ID.

Based on this efficiency profile, we choose a flat 90% tracking efficiency, applied to all smeared, charged particles passing the $p_\perp$ threshold, disregarding any $p_\perp$- and $|\eta|$-dependence for convenience.

As explained in Sec. 1.4, the ATLAS calorimeter system uses topological clustering with a 420-scheme [41], where deposits with energies $E > 4 \times \sigma_{\text{noise}}$ seed the clustering algorithm.

Therefore, based on Fig. 3.6, we use a crude, pseudo-empirical expression for the calorimeter noise as a function of $|\eta|$ given by

$$\sigma_{\text{noise}}(\eta) = \left(0.1 + 0.02\, e^{|\eta|} \right) \text{GeV} \qquad (3.2)$$

Thus, all final state particles (neutral and charged; except muons and

[3] Actually by twice the HCAL resolution to better match the results found in official simulation. Either way, the difference is small and only works to our "disadvantage".
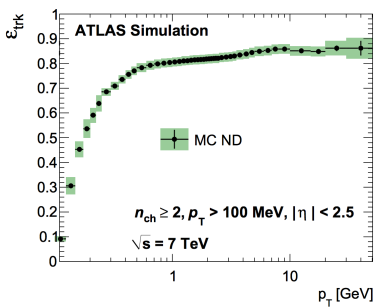

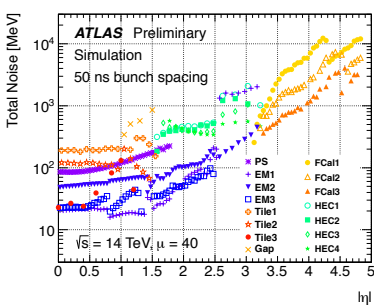Figure 3.5: ATLAS ID tracking efficiency in a $\sqrt{s} = 7\,\text{GeV}$ minbias MC sample. From [28].

Figure 3.6: Total noise in the ATLAS hadronic (tile) calorimeter as a function of $|\eta|$ for the different detector components.
From http://twiki.cern.ch/twiki/bin/view/AtlasPublic/TileCaloPublicResults.

neutrinos) with a HCAL-smeared energy above four times the noise in Eq. (3.2) are reconstructed as TopoClusters with unit efficiency.

We do not correct the three-momentum of charged particles in order to take into account the deflection in ϕ caused by the magnet system. This means that the calorimeter deposits corresponding to charged particles will have the same $y - \phi$ coordinates as the charged particle's track (if reconstructed). There are two reasons for this:

1. The deflection of charged particles in the magnetic field is rather complicated and, what is more, it is $p_\perp$-dependent; therefore, and implementation would be roughly equivalent to implementing a full particle flow algorithm.

2. More importantly, the track based wavelet methods rely on the assumption of a high degree of correlation, in wavelet space, between charged and neutral particles form the same vertex. This assumption is immediately violated by the bending of the charged particles, meaning that—at least to some degree—the track-based methods would be ineffective when applied to standard calorimeter-level input. However, our reluctance to deflect charged particles is not completely based on wishful thinking. As remarked in Sec. 1.4, the particle flow algorithm exactly corrects calorimeter-level objects based on information from tracks with $p_\perp$ up to 40 GeV, after which the deflection of charged particles will be negligible anyway. Therefore, we will be considering our calorimeter objects as particle flow-equivalents in the following.

We also want to make a brief remark on which smearing/inefficiencies we are *not* considering. We already mentioned the deflection of charged particles. More generally, no position smearing in $y - \phi$ is performed for any class of particles; no "crack"-effects are considered; any $p_\perp$ and $|\eta|$ dependence in smearing/thresholds, beyond that in Eq. (3.2), is ignored. Most importantly, no track-vertex matching efficiency is used in the nominal setup. That is, we assume that we are able to correctly associate reconstructed tracks to their production vertex with unit probability. This point is quite central to the performance of the track-based wavelet methods, as they rely on our ability to separate hard scatter tracks from pile-up tracks. As we will see in Sec. 4.1 this assumption is too optimistic, but for the time being we will postpone this discussion, and instead consider the following results a best-case-scenario as far as tracking is concerned.

Finally, we stress that the point of this "toy model" smearing is not to faithfully replicate the actual performance of the different ATLAS detector sub-systems. Rather, we want to test our ideas in a setting which is *roughly* similar to that of LHC general purpose detectors. If these ideas do not come to fruition in this simplified setting, we will have little hope of being able to implement them in official reconstruction software. However, if things do work out here, we will have gained a better understanding of our techniques and will be able to advance them into the more complicated setting of official ATLAS simulation.

■ Wavelet and jet setup

Now we only need to describe our wavelet analysis setup as well as the jet clustering we will be using, before the actual study can commence.

In the following, we have chosen to use the Haar wavelet basis, a rasterisation rapidity range of $|y| < 3.2$, and a 64×64 pixel definition[4]. We choose the Haar basis since it is comparatively simple, cf. Fig. 2.3, since its action on the event is intuitive, and since more complicate bases have not proven to yield improvements in performance[5], cf. Sec. 3.4. We use $|y| < 3.2$, since it allows for a coverage of the entire ATLAS tracking system as well as the entire ATLAS ECAL and the barrel and end-cap sections of the HCAL, cf. Tab. 1.4, meaning that we can use the ATLAS ID for the track-based wavelet methods while at the same time taking full advantage of the central ATLAS calorimeter system, allowing for unhindered jet reconstruction. Also, having almost identical ranges in y and ϕ (6.4 and 2π) ensures that we will have almost square pixels in $y - \phi$. We choose a 64×64 pixel grid, since the resulting resolution in $y - \phi$ is roughly $6.4/64 \times 2\pi/64 \approx 0.1 \times 0.1$, which is similar to the ATLAS HCAL resolution [32], see Sec. 1.3. One could perhaps have been optimistic, and argued for a 128×128 resolution, corresponding roughly to the ATLAS ECAL resolution; however, we choose to be conservative.

Jet clustering is performed using the FASTJET package [77]. We choose to use the Cambridge-Aachen (C/A) clustering algorithm, cf. Sec. 1.4, with a distance parameter of $R = 1.2$, which was the standard in most Run 1 boosted boson analyses [14, 15, 16]. Furthermore, we realise that wavelet methods do not exist in a "vacuum": as mentioned in Sec. 1.4, several techniques already exist for countering the effects of pile-up on jet energy measurements ('pile-up mitigation'; area subtraction, NpC, SoftKiller, PuPPI) and for removing soft contributions to boosted jets ('grooming'; BDRS, trimming, pruning). Most, of not all, of these methods have a proven record of excellent performance and so we must view wavelet analyses in this broader scope of standard approaches, and consider the combined performance of wavelet-based methods with other techniques, where applicable. Our hope is then that, using wavelet-based techniques, we will be able to rise above and beyond what can be achieved by existing means.

As reference methods we will be using the area subtraction method, cf. Sec. 1.4, as a stand-alone pile-up mitigation alternative to wavelet-based pile-up subtraction, and the modified mass-drop filtering technique BDRS-A, cf. Sec. 1.4, both as stand-alone and in conjunction with a wavelet analysis (i.e. on jets clustered in wavelet-cleaned events), as these are some of the most widely used standard methods.

[4] Fig. 2.3 shows the complete set of Haar basis functions for an 8×8 grid. For large grids, additional basis functions continue down to the right of Fig. 2.3 in a similar dyadic pattern.

[5] Additionally, the Haar wavelet is the only one of the Daubechies wavelet which does not suffer under the periodisation requirement mentioned in Sec. 2.1. All other $D_{N>2}$ wavelets will have basis functions which connect regions at $y \approx y_{max}$ with regions at $y \approx -y_{max}$, which might be problematic.

3.2   Wavelet coefficient distributions

Having proposed a small selection of possible wavelet methods, and described the simulated processes in which we want to study their effectiveness, we now see how these methods behave. First, we will study the distributions of the absolute values of coefficients to assess the effect of the geometric methods (FDN and SDN). Second, we will look at the behaviour of the ratio of same-indiced the coefficients in wavelet event from tracks, in order to study the justification and effect of the track-based methods (TS and TF). Finally, we will pass judgment on each of the methods and settle on a wavelet analysis setup for the remainder of this study.

◼   Geometry-based methods

We now want to put our ideas to the test. Our first task is studying the wavelet coefficients themselves and try to understand their behaviour. This is done both to get a better understanding of the space of wavelet coefficients, to which we apply out methods, as well as to see whether the methods proposed in Sec. 2.3 are indeed applicable.

As detailed in Sec. 3.1, we will study $W + $ jets events, generated with PYTHIA8 at $\sqrt{s} = 13\,\text{TeV}$, overlaid with an average number of $\langle \mu \rangle = 40$ minbias collisions (pile-up). Similarly, we use the Haar wavelet on a 64×64 pixel array, extending out to $|y| < 3.2$. In this section, we only study events with a reconstructed jet $p_\perp > 300\,\text{GeV}$ requirement. The distribution of absolute values of the wavelet coefficients, from performing the forward 2D DWT of such rasterised events, are shown in Fig. 3.7 for $(m_y, m_\phi) = (3,3)$.
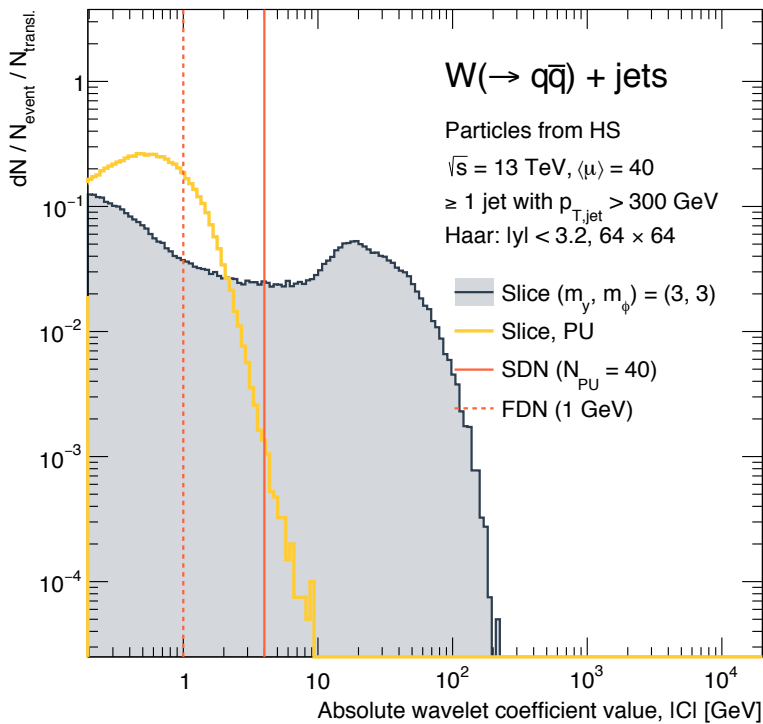
Figure 3.7: The distribution of absolute values of the wavelet coefficients in $W + $ jets samples at $\sqrt{s} = 13\,\text{TeV}$, with at least one reconstructed $\text{C/A}^{R=1.2}$ jets with $p_\perp > 300\,\text{GeV}$ for $(m_y, m_\phi) = (3,3)$. Also drawn is the corresponding distribution in pure pile-up events with $\langle \mu \rangle = 40$ (full yellow) as well as typical SDN ($N_{\text{PU}} = 40$; full red) and FDN (1 GeV; dashed red) cuts.

The scales $m_{y,\phi}$ are chosen to be fairly generic/general. Here, the filled graph shows the distribution of absolute coefficient values in the hard scatter (HS) interaction *alone*; the yellow line shows the distribution in pure pile-up (PU); the full red line shows the scale-dependent denoise (SDN) cut as this set of scales, cf. Eq. (2.45); and the dashed red line shows an indicative flat de-noise (FDN) cut. Each bin is normalised by the number of events and the number of translation indies at $(m_y, m_\phi) = (3, 3)$ (see below)—but *not* by bin area. Fig. 3.7 therefore shows a count-distribution rather than a probability density.

From this plot it is seen that the distribution in pile-up has no real structure, while the hard scatter clearly exhibit a ridge-structure reflective of the high-$p_\perp$ particles. These two distributions are what we want to separate in a way, such that the distribution of modified coefficients in the *full* event resembles the distribution of the hard scatter coefficients as closely as possible.

In order to study the behaviour of the coefficients *across* frequency scales $m_{y,\phi}$, we can plot distributions such as those in Fig. 3.7 for all scales simultaneously. This is done in Fig. 3.8 for pure pile-up.

Figure 3.8: The distribution of absolute values of the wavelet coefficients in pure pile-up events with $\langle\mu\rangle = 40$ at $\sqrt{s} = 13\,\text{TeV}$. Each vertical slice is normalised by the number of events times the number of translations at that particular set of scales, i.e. $2^{\max(m_y-1,0)} \times 2^{\max(m_\phi-1,0)}$. The scale-dependent ($N_{\text{PU}} = 40$; full red) and flat (1 GeV; dashed red) de-noise thresholds are shown together with the RMS (full yellow) of each vertical slice.
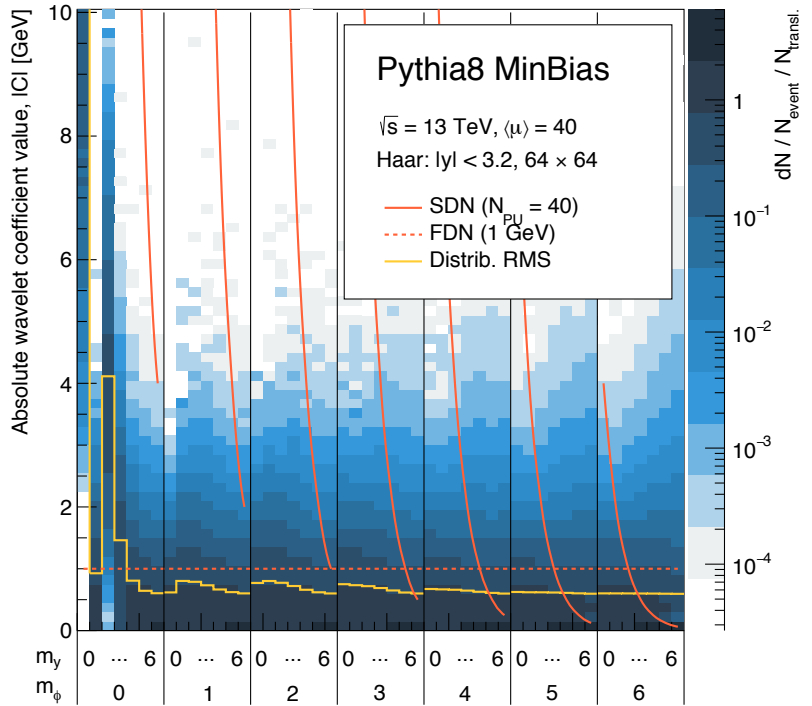
In this plot, each vertical slice corresponds to a distribution like the (yellow) one in Fig. 3.7, for a particular set of wavelet basis function scales, (m_y, m_ϕ), see Fig. 2.3. That is, the first slice contains the coefficients of the basis function with $(m_y, m_\phi) = (0, 0)$ for all events, normalised by the number of events; similar for the second slice; the third slice contains the coefficients of the *two* basis functions with $(m_y, m_\phi) = (2, 0)$, see Fig. 2.3, for all events, normalised by the number of events *and* the number of translations at this set of scales,

which in this case is two. In general, at scales (m_y, m_ϕ) there will be $2^{\max(m_y-1,0)} \times 2^{\max(m_\phi-1,0)}$ translations of the same fundamental basis functions, cf. Sec. 2.1 and see Fig. 2.3, and each vertical slice is normalised by this number to give a per event distribution for each fundamental basis function (regardless of its scale). We have plotted the root-mean-square ('RMS' for short) of the distributions in each vertical bin, to allow us to study the scale-dependent structure of the coefficients in the simulated events. Also shown in the plot are the scale-dependent thresholds of Eq. (2.45), where the average number of pile-up collisions $\langle \mu \rangle$ is used as $N_{\mathrm{PU}} = \langle \mu \rangle = 40$ for clarity. In principle this number will be dependent on the actual number of collisions in each event, and the full red line therefore represents only the average positions of these cuts, which is, however, sufficient to study their general behaviour. We clearly see the purely geometric, scale-dependent nature of these cuts: coefficients with small $m_{y,\phi}$ are all but discarded, while the coefficients are large are left almost intact. Finally, a flat de-noise threshold of $1\,\mathrm{GeV}$ is shown, and we see that it matches well with the flat-baseline structure of the RMS'es at high $m_{y,\phi}$.

The distributions exhibit a number of distinct features, as evidenced by their respective RMS'es, which we will briefly address:

- In the $(m_y, m_\phi) = (0,0)$ slice we see the uniqueness of lowest-scale basis function: it is the only one with a distribution which is *not* symmetric around 0, but instead around $\sim 25\,\mathrm{GeV}$, never being negative. This justifies the special care taken in the track filtering methods, Sec. 2.3.

- The slices with $(m_y, m_\phi) = (1,0)$ and $(0,1)$ represent geometrically identical basis functions, separated only by a rotation of $\pi/2$. However, as is evident from the figure, the basis functions with $(m_y, m_\phi) = (1,0)$ have, on average, an energy/information content which is almost a factor of two greater than those with $(m_y, m_\phi) = (0,1)$. We believe that this is can be attributed to the fact that $(0,1)$ measures the energy asymmetry in the transverse plane while $(1,0)$ measures the energy asymmetry along the beam direction, see Fig. 2.3. The fact that we expect to have (approximate) energy conservation in the transverse plane, but not necessarily in the longitudinal direction, where the initial-state parton momentum fractions $x_{1,2}$ may very well differ significantly, cf. Sec. 1.2, can explain this discrepancy.

- The slices with $(m_y, m_\phi) = (2,0)$ and $(3,0)$ have very large RMS'es compared to the general structure in the remaining slices. In particular, the slice with $(2,0)$ does not peak at zero in contrast to all the other slices (except the first, but still being *symmetric* around 0). We attribute this to the fact that that basis functions with $(m_y, m_\phi) = (2,0)$ and $(3,0)$ are sensitive to variations, along the beam line, which are more high-frequent than just left-right asymmetry. Especially $(2,0)$ is sensitive to the rapidity plateau structure mentioned in Sec. 1.2, which to some, but lesser, extent

also holds for $(3, 0)$. Since we on average expect there to be a significant asymmetry between the energy density at small $|y|$ compared to large $|y|$, it makes sense that exactly these two sets of scales reflects this expectations.

- Finally, we see that the large-scale correlations vanish around $m_{y,\phi} \gtrsim 4$, where the RMS'es flatten out completely.

The distributions in Fig. 3.8 are the ones we want to discard, representing the contributions from pile-up alone, and so we want our cuts to behave in a way which should act to remove this part, while keeping the hard scatter contribution close to intact. In order to asses the effect of the geometrical cuts on the hard scatter coefficients, Fig. 3.9 shows the similar distributions of wavelet coefficients resulting from the FWT of the hard scatter event alone.

Figure 3.9: The distribution of absolute values of the wavelet coefficients in W + jets samples at $\sqrt{s} = 13$ TeV, with at least one reconstructed $C/A^{R=1.2}$ jets with $p_\perp > 300$ GeV. The normalisation, cuts, and RMS dsitributions are the same as for Fig. 3.8, here including also the RMS'es for the pileup only coefficients (dashed yellow) with $\langle \mu \rangle = 40$ as a reference.
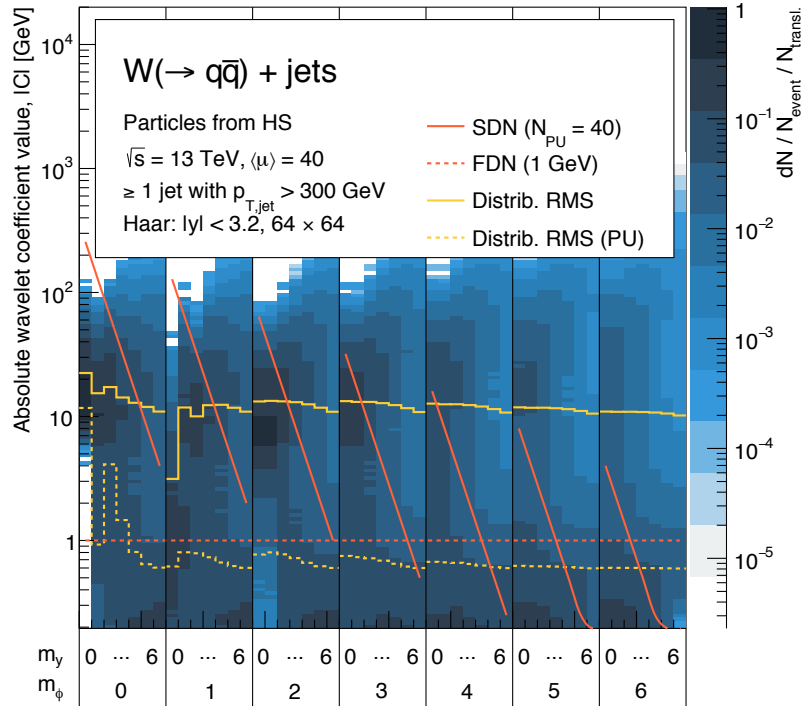
Here, we see many of the same features as in Fig. 3.8, such as the uniqueness of the $(m_y, m_\phi) = (0, 0)$ basis function, the conservation of energy in the transverse plane as measured by the $(0, 1)$ basis function[6], etc. However, we may also address a few features particular to the distributions of hard scatter coefficients:

[6] This features is more pronounced here, since the coefficients in Fig. 3.9 are those for single collision events—because the total uncertainty on the energy conservation in the transverse plane grows with the square root of the number of independent collisions—and each of these have much higher energies than the minbias collisions in Fig. 3.8 individually.

- There is a clear dip-structure for the coefficients with $(m_y, m_\phi) = (1 - 3, 1)$, which we may explain as follows: We see from Eq. (1.32) that the maximal system rapidity for a two-jet configuration with $p_\perp > 300$ GeV is $y_{max} \approx 1.6$. Since the two jets will be rather central and in opposite hemispheres (back-to-back, due to momentum conservation), the $(1, 1)$ coefficient will pick up large contributions

in cases where y_{jet_1} and y_{jet_2} have opposite signs. The basis functions at scales $(2, 1)$ are *centered* around y_{max}, which means that they also only get large contributions in cases where y_{jet_1} and y_{jet_2} have opposite signs. However, these contributions are arguably better captured[7] by e.g. the coefficients with $(1, 1)$, meaning that the $(2, 1)$ coefficients will be suppressed. Finally, the basis functions at scales $(3, 1)$ have widths $\Delta y \approx y_{\text{max}}$ and are therefore able to detect energy patterns of size $\Delta y / 2 \approx y_{\text{max}} / 2$ which is the expect order of the jet-jet separation in y. Therefore, this coefficient might also pick up large contributions in cases where a basis function falls between the two hard jets, or just close to either..

- This last feature is generally seen for all basis functions with $m_y + m_\phi \approx 2 - 4$, since their area in $y - \phi$ is given by

$$A_{m_y, m_\phi} = \frac{2\pi \times 6.4}{2^{\max(m_y - 1, 0)} 2^{\max(m_\phi - 1, 0)}} \approx 4 \times \pi \qquad (3.3)$$

appropriate[8] for detecting energy clusters with characteristic size πR^2 with $R \sim \mathcal{O}(1)$.

- The generally larger energy in the hard event is manifested by a baseline, for $m_{y,\phi} \gtrsim 4$, which is roughly a factor of 20 large than that in Fig. 3.8.

- Finally, we can see a general ridge-like structure, which we recognise from Fig. 3.7. This is a consequence—at small frequency scales—of the localised high energy density deposits, which we cluster to jets, and—at large frequency scales—of single very high-$p_\perp$ particles standing out from their softer counterparts at high.

We have now studied the distributions of wavelet coefficients in simulation, both for pure pileup as well as for high-$p_\perp$ $W + $ jets events, and looked at the behaviour of the purely geometric methods proposed in Sec. 2.3. However, distributions such as those in Figs. 3.8 and 3.9 will not help us to study the validity of the assumptions behind Eqs. (2.46) and (2.47) and the resulting effect of the track-based methods. This is what we turn to now.

■ Track-based methods

In order to study the validity of the track-based methods proposed in Sec. 2.3, we must first test the underlying assumption that tracks alone encode are representative of the event to which they belong as a whole; in particular, that this is true in the space of wavelet functions.

Therefore, we use the fact that we have generator-level ("truth") information available, and study the content of Eq. (2.46). This is done for $(m_y, m_\phi) = (0, 0)$ in Fig. 3.10, which is a correlation plot of the two ratios in Eq. (2.46).

Here we see a strong correlation, however shifted upwards, towards larger values of $|C^{\text{HS}-\text{track}}| / |C^{\text{FE}-\text{track}}|$. This can be explained by the fact that the smearing procedure discards all tracks with $p_\perp < 400\,\text{GeV}$. This mainly affects the contribution from pile-up, having

[7] Since large values for the $(2, 1)$ coefficients imply large energy deposits in the range $1.6 < |y| < 3.2$, which is *not* expected. These coefficients presumably act mainly to suppress the similar effect for the $(1, 1)$ basis function.

[8] The factor of 4 comes from the fourfold division of the 2D Haar wavelet functions.

Figure 3.10: The ratio of the wavelet coefficients at scales $(m_y, m_\phi) = (0,0)$ from the hard scatter alone to those from the full event, for tracks versus that for generator-level particles ("truth"). Plotted for W + jets events with $\langle \mu \rangle = 40$ at $\sqrt{s} = 13$ TeV, with at least one C/A$^{R=1.2}$ jet with $p_\perp > 300$ GeV.

$W(\to q\bar{q})$ + jets

$\sqrt{s}$ = 13 TeV, $\langle \mu \rangle$ = 40
$\geq$ 1 jet with $p_{T,jet}$ > 300 GeV
Haar: $|y| < 3.2$, 64×64
$(m_y, m_\phi) = (0, 0)$

[9] Notice, however, the steep logarithmic scale. This means that the below remarks only concerns a small fraction of the total picture.
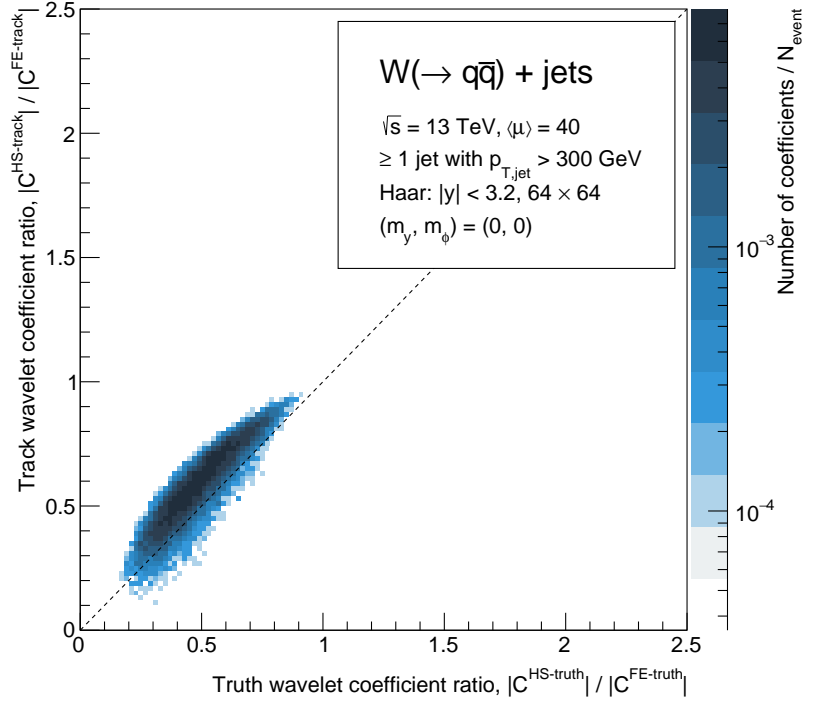
a very soft $p_\perp$-spectrum, cf. Fig. 2.5, and thus shifting the ratio of average energies in the event, i.e. $|C^{\mathrm{HS-track}}|/|C^{\mathrm{FE-track}}|$, upwards. So, at $(m_y, m_\phi) = (0,0)$ the simulation behaves as expected. However, as remarked before, the basis function at $(m_y, m_\phi) = (0,0)$ is unique, so we have to look at other scales as well, to see whether our ideas truly hold up to the test of (simulated) reality. This is done in Fig. 3.11 for $(m_y, m_\phi) = (3,3)$ and $(6,6)$.

In Fig. 3.11(a) a correlation is still clear[9], but considerably more "washed out" than in Fig. 3.10. This can be understood by the fact, that the scales $(m_y, m_\phi) = (3,3)$ are intermediate: that is, they do not capture the features of *total* the event, as in Fig. 3.10, but on the other hand the basis functions still have sufficiently large support to capture moderate-scale energy patterns like jets, see the above remarks. The prominent peak at coordinates $\approx (1,1)$ corresponds to cases where a number of high-$p_\perp$ particles from the hard scatter event carry almost the entire energy in the coefficient (e.g. close to, or in, a jet). The rise towards coordinates $(0,0)$ similarly corresponds to cases where the basis functions are in the space *outside* the jets, where only UE activity from the hard interaction is expected. We also see that we occasionally have hard scatter-to-full event coefficient ratios which are greater than one. Since the hard scatter objects (either tracks or truth particles) form a subset of objects in the full event, this might seem odd at first sight. However, since we are talking about coefficients akin to those in a Fourier expansion, and not e.g. pixel values, the hard scatter coefficients might indeed have larger values than those in the full event. For instance, if some number of pile-up particles fall

just slightly away from particles from the hard scatter, since wavelets measure *differences* in energies, a wavelet basis function for the hard scatter alone might see a large difference and consequently pick up a large contribution, whereas one for the full event will see a smaller difference and thus have a smaller value. This serves to remind us that wavelet coefficients (as measures of energy differences) do not simply add, unlike pixel $p_\perp$'s (as measures of total energy depositions).

(a) At scales $(m_y, m_\phi) = (3, 3)$.
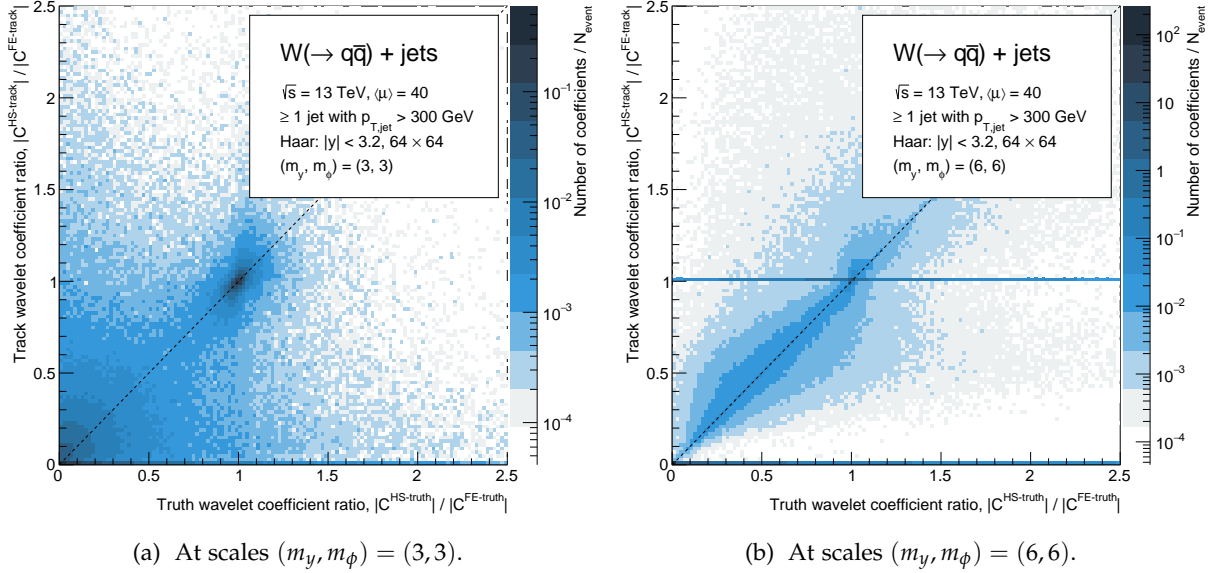
(b) At scales $(m_y, m_\phi) = (6, 6)$.

Fig. 3.11(b) shows a number of striking features:

- A vast fraction of the coefficients fall in the bin with coordinates $(0, 0)$, corresponding to areas of the detector where no (hard scatter) particles fall. This features is particularly evident at the smallest scales, where the wavelet function support is minimal, cf. Eq. (3.3), and therefore the probability for a particle to hit within a particular $\sim 0.2 \times 0.2$ segment in $y - \phi$ is vanishing.

- The horizontal line with $|C^{\text{HS}-\text{track}}|/|C^{\text{FE}-\text{track}}| = 0$ corresponds to the case where a truth-level hard scatter particle (neutral or charged) hits within the wavelet support, but is not reconstructed as a track. This region is mostly thought to be occupied by neutral particles, which naturally never will be identified as tracks, and to a smaller extent by very soft or non-reconstructed charged particles.

- The horizontal line with $|C^{\text{HS}-\text{track}}|/|C^{\text{FE}-\text{track}}| = 1$ corresponds to situations here exactly one reconstructed track—from the hard scatter interaction—falls within the support of a particular basis function. This horizontal slice is peaked at $|C^{\text{HS}-\text{truth}}|/|C^{\text{FE}-\text{truth}}| = 1$ and then falls off on both sides, which can be attributed to collinear neutral particles as well as soft, not-reconstructed activity around the track.

- Finally, the remainder of the plots covers the cases, where more than just a single hard scatter tracks is reconstructed within the

Figure 3.11: The ratio of the wavelet coefficients at scales $(m_y, m_\phi) = (3, 3)$ and $(6, 6)$ from the hard scatter alone to those from the full event, for tracks versus that for generator-level particles ("truth"). Plotted for $W + $ jets events with $\langle \mu \rangle = 40$ at $\sqrt{s} = 13\,\text{TeV}$, with at least one $C/A^{R=1.2}$ jet with $p_\perp > 300\,\text{GeV}$.

wavelet function support, yielding a smeared, yet considerable correlated picture of the relation in Eq. (2.46).

Generally, we see that, from the largest to the smallest scales, we do in fact observe a relation of the form proposed in Eq. (2.46), however, with varying degrees of correlation.

In Fig. 3.12 we show the scale-wise distributions of $|C^{\mathrm{HS-track}}|/|C^{\mathrm{FE-track}}|$ ratios in $W+$ jets events overlaid with $\langle\mu\rangle = 40$ pile-up events. This is done to assess the impact of the track-filtering method across all scales. Each vertical slice then corresponds to projection onto the y-axis of plots like those in Figs. 3.10 and 3.11.

Figure 3.12: The distribution of ratios of same-indiced wavelet coefficients, for hard scatter tracks only to those for tracks in the full event, in $W+$ jets events. A typical track filtering threshold is also shown (0.7; dashed red).
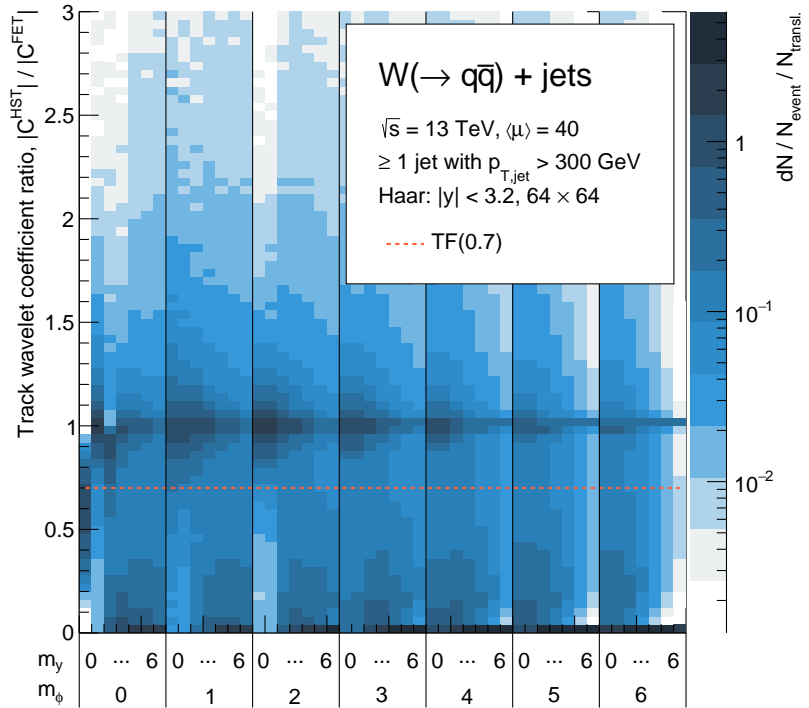
Here we notice a few trends:

- Again, the basis function at scales $(m_y, m_\phi) = (0, 0)$ plays a special role, and we see why it made sense to make an exception in Fig. 2.9: we would be killing the average energy component in the majority of events.

- The distribution for basis functions at $(2, 0)$ has a peak at $0.8 - 0.9$ and then a sharp drop at 1. We recall, that these are the basis functions which are sensitive to large scale differences along the beam direction, in particular to the edge of the ATLAS ID at $|\eta| = 2.5$. Since the basis functions extend out to $|y| = 3.2$ and are centered around $y = 1.6$, this means, that if $\sum_{|y_{\mathrm{trk}}|<1.6} p_\perp \gg \sum_{|y_{\mathrm{trk}}|>1.6} p_\perp$, as will mostly be the case due to the rapidity plateau structure *as well as* the ID edge, these coefficients measure roughly the fraction of the energy in the full event tracks within $|y| < 1.6$ attributable to the hard scatter tracks. This will never by exactly

one, due to the presence of pile-up hence the sharp drop. However, in the more involved cases with $\sum_{|y_{\mathrm{trk}}|<1.6} p_\perp \sim \sum_{|y_{\mathrm{trk}}|>1.6} p_\perp$, this reasoning doesn't hold since contributions from individual tracks don't just add, yielding a smoothly falling distribution covering the full range in $|C^{\mathrm{HST}}|/|C^{\mathrm{FET}}|$.

- Apart from the two bins just mentioned, all other scales show a very prominent peak at $|C^{\mathrm{HST}}|/|C^{\mathrm{FET}}| = 1$, giving us further confidence in Eq. (2.46) across all scales.

- Finally, large fractions of the coefficients have a ratio of zero, corresponding to cases with no hard track contributions. This is particularly true at small scale, which should be expected due to their increasingly small extent in $y - \phi$, as noted above.

We believe that we understand the general structure of Fig. 3.12 across all scales, the predominant feature being sharp peaks at one (which narrow at smaller scales) set on more or less smoothly falling backgrounds, with a spike at 0 from coefficients with no hard scatter contributions. This structure is illustrated by the vertical slice for $(m_y, m_\phi) = (3, 3)$ in Fig. 3.13.
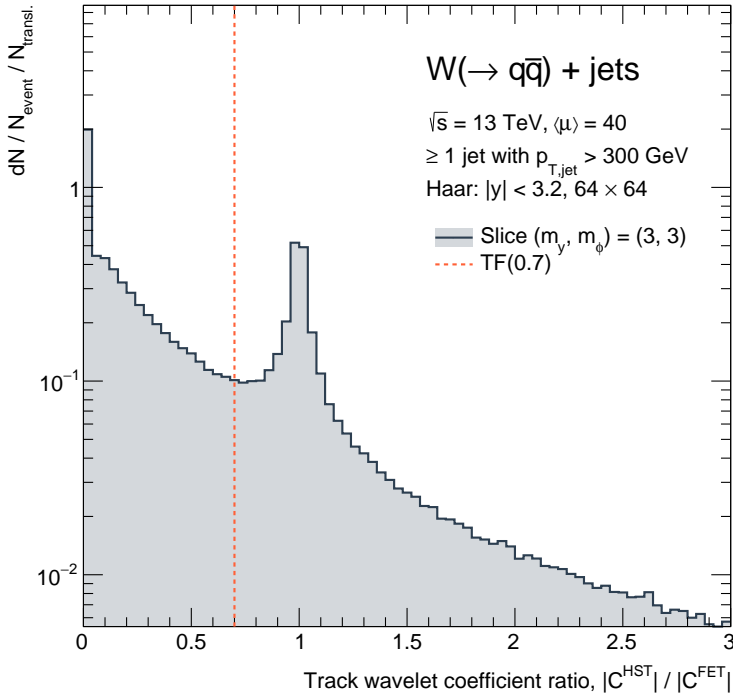
Figure 3.13: The distribution of ratios of same-indiced wavelet coefficients, for the case of scale indices $(m_y, m_\phi) = (3, 3)$, for hard scatter tracks only to those for tracks in the full event, in $W + $ jets events. A typical track filtering threshold also shown (0.7; dashed red).

The peaks seen at all scales in Fig. 3.12, all have small but finite widths, as seen in Fig. 3.13. These can be understood as cases where a basis function is completely dominated by single/few high-$p_\perp$ hard scatter tracks, and any (softer) pile-up tracks only serve to shift the ratio slightly above or below one. This is schematically illustrated in Fig. 3.14.

$$|C^{\mathrm{HST}}/C^{\mathrm{FET}}| = 1$$

(a)

$$|C^{\mathrm{HST}}/C^{\mathrm{FET}}| \gtrsim 1$$

(b)

$$|C^{\mathrm{HST}}/C^{\mathrm{FET}}| \lesssim 1$$

(c)

Figure 3.14: The effect of adding a single soft pile-up track (light blue) to a single hard scatter track (dark blue) on the wavelet basis function (red). The dashed function indicates the wavelet coefficient level in the absence of pile-up tracks.

Similarly the continuously falling background can be understood as the cases with several and/or equally soft tracks, where the addition of one additional track can shift the ratio towards very small or very large values, and even flip the sign of the coefficient. This is schematically illustrated in Fig. 3.15. This smooth background, having a power-law–like structure may be related to the steep $p_\perp$-spectrum of individual pile-up particles, cf. Fig. 2.5.

■ Our choice of methods

We believe that we have an understanding of the behaviour of the wavelet coefficients in pure minbias events as well as in hard scatter interactions, cf. Figs. 3.7, 3.8, and 3.9. Similarly, we are fairly confident in our understanding of the co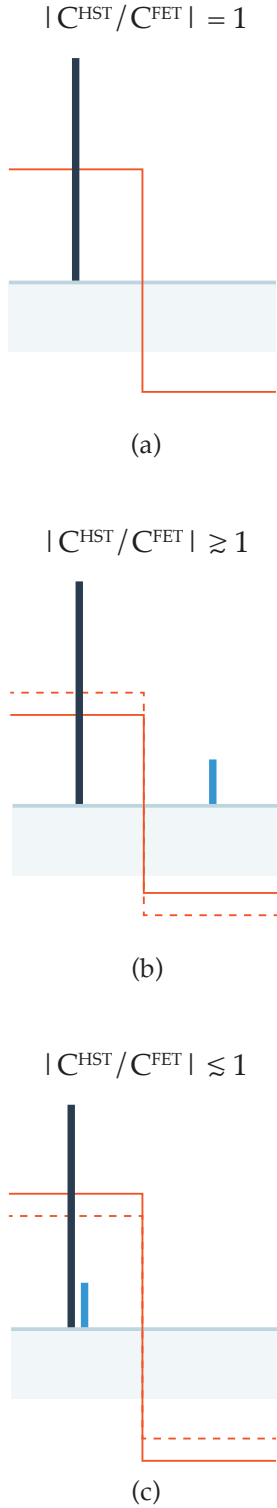rrelation proposed in Eq. (2.46), as studied in Figs. 3.10 and 3.11, as well as the $|C^{\mathrm{HST}}|/|C^{\mathrm{FET}}|$ distributions at all scales, cf. Figs. 3.12 and 3.13. Therefore, we can turn to discussing the wavelet methods suggested in Sec. 2.3.

Flat de-noising (FDN)s. The FDN methods, with a threshold of $\mathcal{O}(1\,\mathrm{GeV})$ seems quite reasonably, cf. Fig. 3.9: it captures well the constant baseline-structure at large $m_{y,\phi}$, while also roughly matching the general level of the RMS'es in pile-up for small $m_{y,\phi}$. The flat threshold discards the coefficients with the lowest information content while being considerable below the crucial features of the hard scatter. However, due the simplicity of this method—especially the fact that does not take into account the evident small-$m_{y,\phi}$ correlations—we do not expect any stellar performance.

Scale-dependent de-noising (SDN). The SDN method goes a bit further, and thus has a structure which we may compare to the RMS'es in simulation. Here we see, that the cuts are *extremely* hard at low scales (i.e., below $m_{y,\phi} \lesssim 2-3$), effectively killing all basis functions, both for pile-up *and* for the hard scatter. In turn, the cuts at the smallest scales are equally loose, keeping all but the completely redundant basis functions. As is clear from Fig. 3.9, the behaviour of the SDN cuts is very far removed from that of the coefficients in pile-up. This can be traced to the fact the the wavelets at scales $m_{y,\phi} > 0$ measure energy *differences* rather than energy averages, as implied in [67] and relayed in Sec. 2.3. However, the simple geometric structure of the cuts in Eq. (2.45) has some interesting consequences: killing the soft, low-frequency contributions while keeping almost all of the fine-grained information could prove useful e.g. for the purpose of jet substructure studies. However, in the following, due to its ruthlessness towards the low-frequency basis functions making this method—in its current form—unsuitable for at least jet energy measurement, we will not be considering it any further, except for reference.

Track scaling (TS). Due to the "washed-out" nature of the ratio of tracks coefficients versus that of truth coefficients, particularly evident in Fig. 3.11(a), scaling the the full event coefficient by the

ratio $|C^{\text{HST}}|/|C^{\text{FET}}|$ is not immediately justified. One problem is, that the many off-diagonal cases would be *forcing* the full event coefficients in directions which are not consistent with the underlying "truth" information. Another is, that we would potentially scale some coefficients by factors of $\gtrsim 2$. This is not to say that such a procedure is categorically prohibited, but a scaling of this sort should be performed with great care. However, from Fig. 3.10 we also see that, at least for some scales, a scaling such as the one proposed in the TS method, is indeed possible and well motivated. Therefore, it might be possible to remedy these problems by considering only certain scales $m_{y,\phi}$, certain absolute full event coefficient values, certain regions of $|C^{\text{HST}}|/|C^{\text{FET}}|$, etc. However, in the following, we will consider the correlation too weak to justify the general use of Eq. (2.47).

Track filtering (TF). Nevertheless, we still observe some structure, which we might be able to exploit in a less invasive way, e.g. in using the TF method. This is further corroborated by Figs. 3.12 and 3.13 where a cut at ~ 0.7 seems very reasonable, keeping the prominent peak while discarding the coefficient with less/negligible hard scatter contributions. Also, considering the exception for $(m_y, m_\phi) = (0,0)$ we believe that this method is sufficiently "safe" and well motivated to warrant further study. Since a flat TF threshold is fairly simple, one might go further and use the additional structure of the distributions in Figs. 3.12, in the spirit of the SDN method; use cuts symmetric around $|C^{\text{HST}}|/|C^{\text{FET}}| = 1$; consider fall-back cuts in cases where $|C^{\text{FET}}| = 0$; etc. However, in this first iteration we choose the simplest and most conservative implementation, shown in Fig. 2.9.

Based one the above considerations, and inspired by the general pass-structure suggested in Sec. 2.3, we will be using a two-pass track filtering and particle cleaning (TF $\oplus$ CL) wavelet analysis, intended to first remove the coefficients deemed to be dominated by pile-up, and the to remove the particles which, as a result, are scaled too much to be associated to the hard scatter event. One might also include a third FDN pass, to perform some final cleaning, at the expense of an additional free parameter. For the time being we choose the simple two-pass structure for convenience of optimisation.

$|C^{\text{HST}}/C^{\text{FET}}| = 1$

(a)

$|C^{\text{HST}}/C^{\text{FET}}| \gg 1$
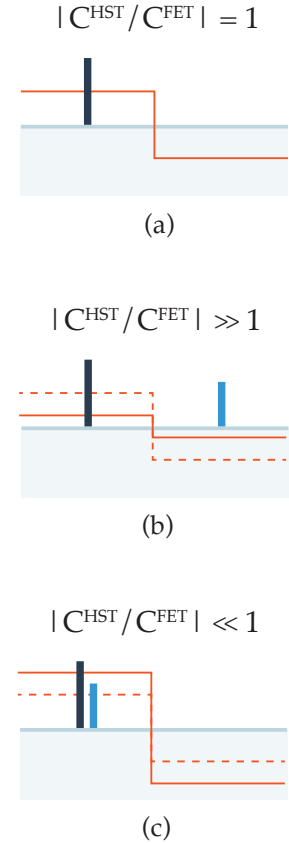
(b)

$|C^{\text{HST}}/C^{\text{FET}}| \ll 1$

(c)

Figure 3.15: The effect of adding a single soft pile-up track (light blue) to a single soft scatter track (dark blue) on the wavelet basis function (red). The dashed function indicates the wavelet coefficient level in the absence of pile-up tracks.

3.3   Optimisation

In the previous section we studied various distributions of wavelet coefficients in toy model simulation and, on this basis, assessed the merits of the various proposed wavelet methods, choosing to use the two-stage TF $\oplus$ CL wavelet analysis to use in the following. In this section, we will use toy model simulation to choose suitable parameter values for both TF and CL at the nominal setup, cf. Sec. 3.1, and check whether this choice is dependent on the type of decaying vector boson, on the boson $p_\perp$, or on the average pile-up multiplicity $\langle \mu \rangle$. Finally, we discuss the impact of the optimised setup on simulation and compare it to the FDN and SDN reference methods.

■   Optimisation procedure

Having settled on a general structure ('track filtering' followed by 'particle cleaning'; TF $\oplus$ CL) for our wavelet analysis, we turn to the determination of an optimal parameter configuration. The parameters for TF and CL are both constrained to the interval $[0, 1]$ (representing fractions) but are otherwise free.

First, we have to decide on a optimisation procedure which is simple, intuitive, and aligned with the metrics presented in Sec. 3.1. As noted there, we are interested in reconstructing the mass of jets originating from boosted vector bosons as accurately (no bias) and precisely (narrow width) as possible, while also possibly discarding QCD jet from the M_jet region around the boson pole masses. A reasonable figure of merit was chosen to be $S/\sqrt{B}$, where we take S to be the number of boson jets within the narrowest M_jet interval containing 68% of the total number of boson jets, and B to be the number of QCD jets within the same interval. To obtain a sample of boson jets (S) we will be using $W +$ jets events, where jets matched within[10] $dR = 0.6$ to a generator-level W boson are labeled as "W-jets". Based on the discussion in Sec. 3.1, we obtain a sample of QCD jets (B) from QCD dijet events.

Also, as mentioned in Sec. 3.1, we need to consider the effect of wavelets analyses in conjunction with existing methods. For jet grooming, we have chosen the modified mass-drop filtering (BDRS-A) as our base of reference, and our figure of merit will be $S/\sqrt{B}$ for the combined TF $\oplus$ CL $\oplus$ BDRS-A setup[11] *relative* to the value for BDRS-A alone. This tells us how much is gained by the wavelet analysis on top of what is achievable by existing means.

Finally, since the BDRS-A method occasionally kills entire jets (yielding $M_\text{jet} \sim 0$, cf. Sec. 3.4), we choose a lower limit of $M_\text{jet} > 20\,\text{GeV}$ in finding the narrowest interval, and require the interval to contain 68% only of jets with $M_\text{jet} > 20\,\text{GeV}$.

Our hope is then, that a parameter configuration optimised on the basis of this figure of merit, specific to the boson mass metric, will also be suitable for the jet energy measurement. This hope is based in the fact that our figure of merit is chosen to measure how faithfully the wavelet analysis is able to reconstruct a certain kinematic quantity (mass) by removing the contribution from pile-up; therefore,

[10] This figure was found, from dR_{JW} distributions in simulation, to be suitable.

[11] That is, using BDRS-A groomed jets clustered from the output collection of particle flow-like objects resulting from the combined TF $\oplus$ CL method, cf. Fig. 2.6.
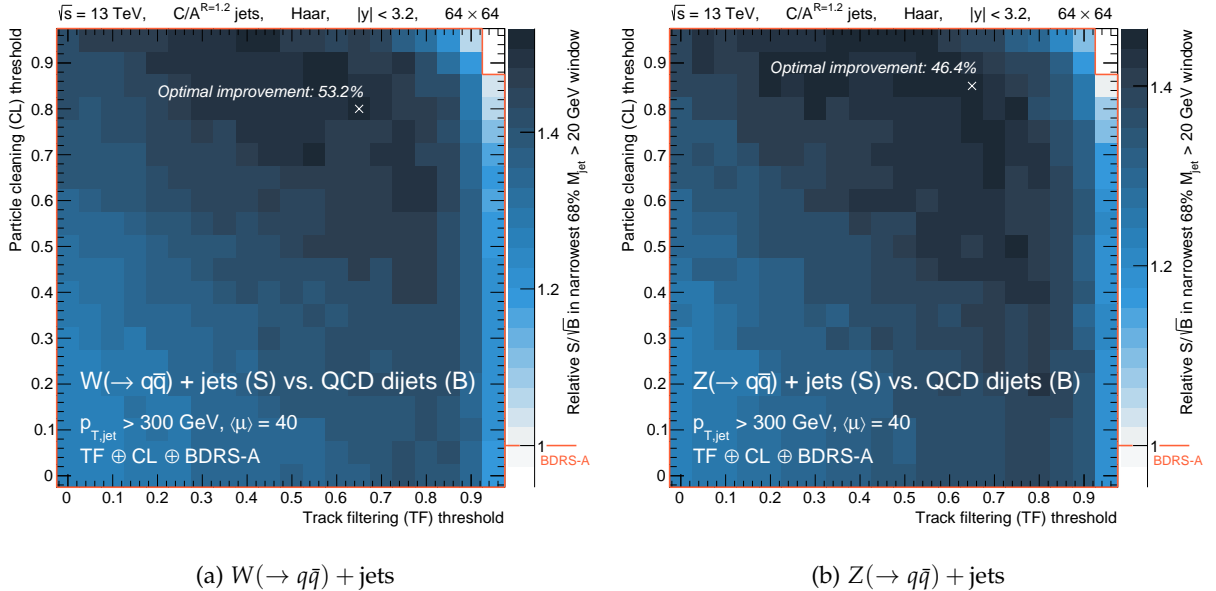
it might very well be that the same configuration is also suitable for reconstructing another, but related, kinematic quantities (energy).

We will mainly be performing the optimisation for our nominal setup: $W + \text{jets}$ events at $\sqrt{s} = 13\,\text{TeV}$ with $\langle \mu \rangle = 40$ PYTHIA8 minbias events, with at least one reconstructed, BDRS-A groomed $\text{C/A}^{R=1.2}$ jet with $p_\perp > 300\,\text{GeV}$, using the Haar wavelet on a 64×64 pixel array with $|y| < 3.2$. All of these choices constitute coordinate axes in a large-dimensional "setup space", which we cannot possible search exhaustively. We therefore assume that the optimal parameter configuration for the nominal setup will be the same for all other configurations. This is a strong assumption, and therefore we also perform optimisations at different points along what we have found to be the decisive coordinate axes: $\langle \mu \rangle$ and $p_{\perp \text{jet}}$. Additionally, a few checks using different setups are performed in Sec. 3.4.

■ Nominal setup

Since we have chosen a wavelet analysis with only two free parameters, we can search the parameter space easily and densely. We do this, by using a regular 20×20 grid in the parameters of TF and CL, resp., with equidistant points along each axis: $[0, 0.05, 0.1, \ldots, 0.90, 0.95]$. We omit coordinates $\text{TF}(1.0)$ and $\text{CL}(1.0)$ since either is too hard and will yield uninteresting results.

For each coordinate point, we compute $S/\sqrt{B}$ for the TF $\oplus$ CL $\oplus$ BDRS-A setup and compare it to that with just BDRS-A grooming. The result is shown in Fig. 3.16, both for W and Z boson jets.

(a) $W(\to q\bar{q}) + \text{jets}$

(b) $Z(\to q\bar{q}) + \text{jets}$

Figure 3.16: Optimisation plots for $W/Z(\to q\bar{q}) + \text{jets}$, with $p_{\perp \text{jet}} > 300\,\text{GeV}$ at $\langle \mu \rangle = 40$. Red line indicates performance for BDRS-A alone.

Each of the point in these plots, and the following, corresponds to a M_jet distributions like the one shown in Sec. 3.4. We briefly discuss the features of both optimisation plots:

• We see a general improvement of up to 50% relative to standard

BDRS-A. More importantly, we see an improvement for practically *all* parameter configurations.

- The optimisation contours follow an arc-like shape, roughly from $(0, 0.9)$ over $(0.7, 0.7)$ to $(0.9, 0)$. This can be understood in the following way: given a low CL threshold, one needs to impose a large TF cut, since then more coefficients will be killed, the pixels in the output rasterised event will change correspondingly more, scaling the momentum of pile-up particles further downwards, and thus making the low CL cut effective. The converse reasoning also hold for low TF/high CL cuts.

- At the highest parameter values (in particular for TF), we se a clear degradation in performance, justifying the omission of the TF(1) and/or CL(1) configurations.

- At the $(0, 0)$ point in parameter space we don't return to the usual BDRS-A performance, as one might initially think, but rather to a level $25 - 30\%$ above this. That is due to the use of the '$\leq$' criteria in the TF method, cf. Fig. 2.9, meaning that coefficients with $C^{\mathrm{FET}} \neq 0$ and $C^{\mathrm{HST}} = 0$ are killed, as remarked in Sec. 2.3. This leads to a minimal cleaning even for the $(0, 0)$ configuration. Indeed, if one were to use a strict inequality in the TF method instead, all points with TF(0) would have unit figure of merit.

- Finally, we see that the structure of the optimisation plots, the optimal parameter configurations, and the performance improvement at these configurations are practically identical for both W and $Z +$ jets events. This implies that our choice of wavelet analysis performs well, and in the same way, regardless of the boson type.

Since the optimisation contours are plateauing around in the upper right part of the plots, and due to symmetry, we choose TF(0.7) $\oplus$ CL(0.7) as our nominal parameter configuration.

■ Additional setups

This parameter configuration has been shown to work well at $\langle \mu \rangle = 40$ and for $p_{\perp\mathrm{jet}} > 300\,\mathrm{GeV}$. However, for us to use it generally, across a range of different conditions, we want to perform some additional optimisation test. First, we want to see if the conclusions drawn above hold true if we increase the jet $p_{\perp\mathrm{jet}}$ cut. This is what is shown for $W +$ jets events in Fig. 3.17 for $p_{\perp\mathrm{jet}} > 500$ and $1000\,\mathrm{GeV}$.

We again identify a few characteristic features:

- We see that the optimal $S/\sqrt{B}$ improvement is lower at $p_{\perp\mathrm{jet}} > 500\,\mathrm{GeV}$ than at $300\,\mathrm{GeV}$, but still considerably large, while having completely disappeared at $p_{\perp\mathrm{jet}} > 1000\,\mathrm{GeV}$. This can be attributed to three facts. First, the energy resolution for HCAL clusters decreases as $\propto 1/\sqrt{E}$, modulo the constant term cf. Tab. 1.4, and, as a result, the energy of the jet will naturally be better resolved in highly boosted topologies. Second, due to the high $p_\perp$ of the jet, the

(a) $p_{\perp\text{jet}} > 500\,\text{GeV}$

(b) $p_{\perp\text{jet}} > 1000\,\text{GeV}$

Figure 3.17: Optimisation plots for $W(\to q\bar{q}) + \text{jets}$, with $p_{\perp\text{jet}} > 500$ and $1000\,\text{GeV}$ at $\langle\mu\rangle = 40$.
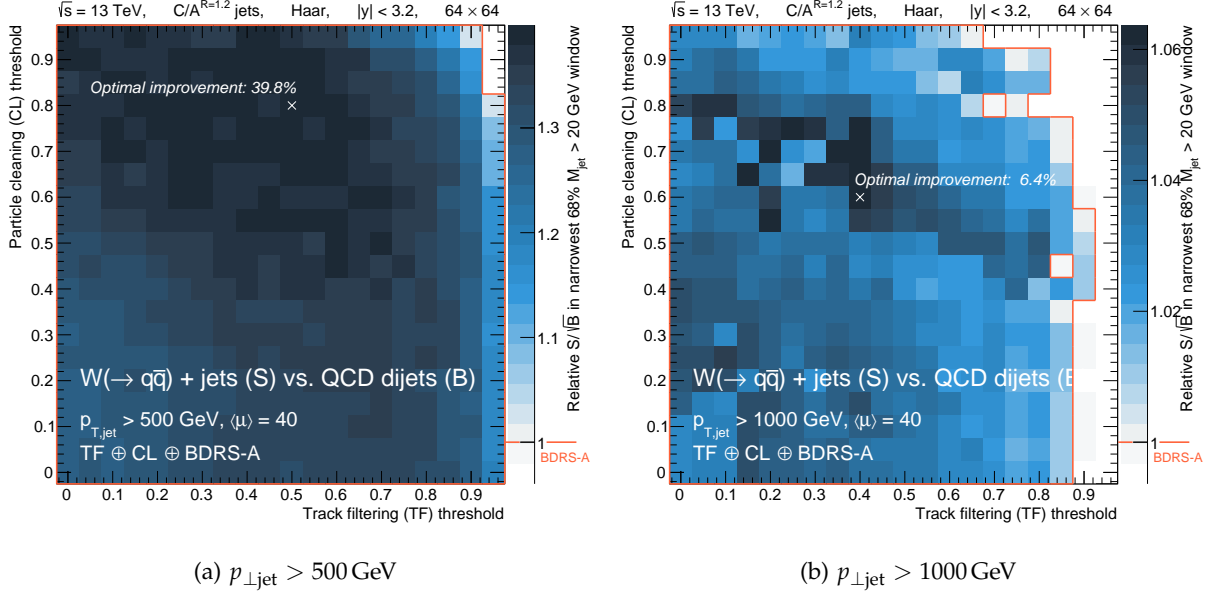
hard constituent particles will be highly collimated, cf. Eq. (1.53), thereby benefiting from the pure dR-weighting of the Cambridge-Aachen algorithm, see Sec. 1.4. This means that any soft, less colli-mated contribution from pile-up and UE will likely be clustered last and is thus more easily removed by the BDRS-A procedure, see also Sec. 1.4. Third and finally, the contribution to the jet's energy from pile-up is on average the same for all optimisation in Figs. 3.16 and 3.17, i.e. $\delta E_{\text{PU}} \approx 0.4\,\text{GeV} \times \langle\mu\rangle \times \pi\,1.2^2 \approx 70\,\text{GeV}$ see Sec. 1.4. At large $p_{\perp\text{jet}}$ this contribution will be relatively smaller, and potential effect of wavelet-based cleaning will be correspond-ingly small. All of these effects serve to establish that BDRS-A filtered C/A jets alone are quite performant under conditions of large $p_{\perp\text{jet}}$, and therefore the potential for improvement decreases.

- We again see a general, arc-like structure in the plots, which, however, washes out at high $p_\perp$. The jaggedness seen in Fig. 3.17(b) is a result of the M_{jet} distribution binning: the narrowest 68% interval might be shifted one bin between adjacent points in the figure and since we are considering already very narrow M_{jet} peaks, such a shift may result in discontinuous jumps in the relative $S/\sqrt{B}$ which hold no physical significance.

- Finally we see that, even under conditions where standard meth-ods fare excellently (very high $p_{\perp\text{jet}}$ and relatively low $\langle\mu\rangle$), the proposed wavelet analysis, for moderate choices of TF and CL pa-rameters, never significantly *deteriorate* performance. Disregarding the rightmost part of Fig. 3.17(b), clustering jets in the wavelet processed event results in considerable improvements at best, and a similar performance at worst.

Having studied the performance of our wavelet analysis versus $p_\perp$ we now turn to its performance under conditions of pile-up multiplicities. Fig. 3.18 shows similar contour plots for $W + \text{jets}$ events with $p_{\perp\text{jet}} > 300$ and $1000\,\text{GeV}$, resp., at $\langle\mu\rangle = 200$.

(a) $p_{\perp\text{jet}} > 300\,\text{GeV}$ · · · (b) $p_{\perp\text{jet}} > 1000\,\text{GeV}$
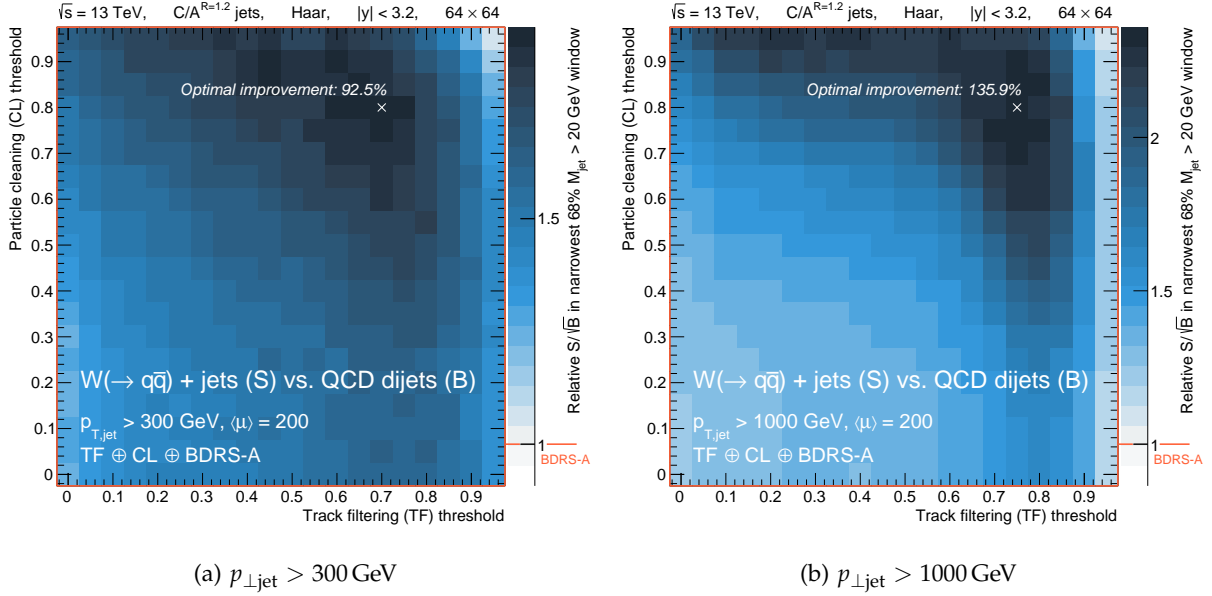
Figure 3.18: Optimisation plots for $W(\to q\bar{q}) + \text{jets}$, with $p_{\perp\text{jet}} > 300$ and $1000\,\text{GeV}$ at $\langle\mu\rangle = 200$.

Being quite similar to the previous optimisation plots, we only want to make a few remarks:

- Also at high pile-up multiplicities the $\text{TF}(0.7) \oplus \text{CL}(0.7)$ parameter configuration appears to be a good choice.

- The improvements in $S/\sqrt{B}$ are considerably larger at $\langle\mu\rangle = 200$ than at 40, cf. Fig. 3.16, confirming our initial hope, that wavelet-based methods could serve as efficient pile-up removal tools.

- Finally, the improvement at $p_{\perp\text{jet}} > 1000\,\text{GeV}$ is now *larger* than at $300\,\text{GeV}$. Our studies suggest that at very high pile-up multiplicities, the BDRS-A method is rendered ineffective: since the filtering is based solely on momentum balance, cf. Sec. 1.4, in situations with low to moderately high $p_\perp$ jets embedded in high pile-up–density environment[12], chance momentum balances of $y > 0.04$ occur before the jet is stripped of its soft constituents[13]. This results in jets that are insufficiently groomed, and therefore M_{jet} values which are far above the boson pole mass. The wavelet analysis improves the resolution at $p_{\perp\text{jet}} > 300\,\text{GeV}$ significantly, but the hard jet is still too soft for the BDRS-A to work optimally, even after pile-up subtraction. However, at sufficiently high $p_{\perp\text{jet}}$ (i.e. $1500 - 2000\,\text{GeV}$) the jet carries enough momentum to stand out to a sufficiently high degree, such that the BDRS-A procedure is not stopped before the hard jet is reached. Since the $p_{\perp\text{jet}} > 1000\,\text{GeV}$ cuts leaves us the the region, where BDRS-A on itself is not yet fully effective, the removal of pile-up by the wavelet analysis creates conditions where the subsequent BDRS-A filtering *is* indeed fully effective, allowing us to see large improvements in our figure of merit.

We are now confident that we understand the behaviour of the optimisation plots in Figs. 3.16, 3.17, and 3.18. This means, that we can have at least some faith in our choice of optimal wavelet method parameters and the resulting improvements found in most cases.

[12] We expect, cf. Sec. 1.4, a pile-up energy density of $\langle\rho\rangle \approx 0.4\,\text{GeV} \times \langle\mu\rangle = 80\,\text{GeV}$ per unit of $\eta \times \phi$, resulting in energy contributions from pile-up of $\delta E_{\text{PU}} \approx \langle\rho\rangle \times \pi R^2 \approx 350\,\text{GeV}$; comparable to the $p_\perp$ of the hard jet.

[13] Note that the BDRS-A method does *not* impose a mass-drop cut.

Note, however, that the figure of merit used in the above only reflects the narrowing of the boson jet mass peak as well as any QCD jet rejection—but not the bias of the jet distribution. For resonant jets, we want M_{jet} distributions which are centered on the pole mass of the respective bosons, but large pile-up multiplicities add energy to the jet which shifts the distribution towards larger values. That is, in addition to the relative improvements upwards of $\sim 100\%$ shown here, our wavelet analysis will reduce the bias of the distributions as we will see in Sec. 3.4.

Finally, we must stress that the comparison with standard grooming methods (here: BDRS-A) is not completely fair either: these methods were conceived and optimised for run conditions with around $\langle \mu \rangle \sim 20$, and so they are not *expected* to perform well at the extreme pile-up multiplicities considered here. But in the face of the results of this sections, it is clear that, even though they were not intended for them, standard filtering methods are rendered ineffective by extreme pile-up conditions, illustrating the need for dedicated pile-up removal tools to remedy this problem. We hope to have shown that wavelet-based methods have the potential to be such a tool.

■ Impact of optimisation on coefficients and particles

We have settled on a nominal wavelet analysis setup of $\mathrm{TF}(0.7) \oplus \mathrm{CL}(0.7)$, and have seen that this choice is reasonable for a range of different values for $\langle \mu \rangle$ and $p_{\perp \mathrm{jet}}$. Both of these cuts are moderately hard, but we have previously seen e.g. in Fig. 3.13 that at least the TF cut seems sensible. Since this parameter choice yields good results for our chosen figure of merit, we can also study its effect on the wavelet coefficients themselves. We cannot show the effect of the TF method as a simple cut, as is possible for the FDN and SDN methods, since the cut is not imposed on the full event coefficients themselves, but rather on the track ratio in Eq. (2.46), which in return gives *modified* coefficient distributions as the one shown in Fig. 3.19.

The plot shows the standard distribution of coefficients in the full events (the starting point), the coefficient of the hard scatter interaction (where we should ideally end), and the coefficients which are kept by the $\mathrm{TF}(0.7)$ method. We see that the TF methods leaves the most energetic coefficients practically unchanged. This is desirable, since these should encode vital information about the hard scatter component of the event, being far more energetic than the pile-up. This can also be seen by the fact that the light blue and yellow graphs converge for $|C| \gtrsim 10\,\mathrm{GeV}$. In the low energy region, with $|C| \sim \mathcal{O}(1)$, the $\mathrm{TF}(0.7)$ method reduces the number of coefficient by almost an order of magnitude, bringing the distribution closer to that of the hard scatter event. This behaviour is also desirable, since we want the method to remove non-vital, soft components from pile-up and UE. But we also notice that this method does no remove *all* of the soft coefficients—which, looking at the yellow graph, also seems unjustified—but instead keep the ones we believe are important to the

Figure 3.19: The distribution of coefficients kept by the TF(0.7) method (filled dark blue) in W + jets events with the nominal setup cf. Sec. 3.1 at scales $(m_y, m_\phi) = (3, 3)$. Also drawn is the corresponding distribution in the full, unmodified event (full light blue) and the in the hard scatter event alone (full yellow) as well as typical SDN ($N_{PU} = 40$; full red) and FDN (1 GeV; dashed red) cuts.
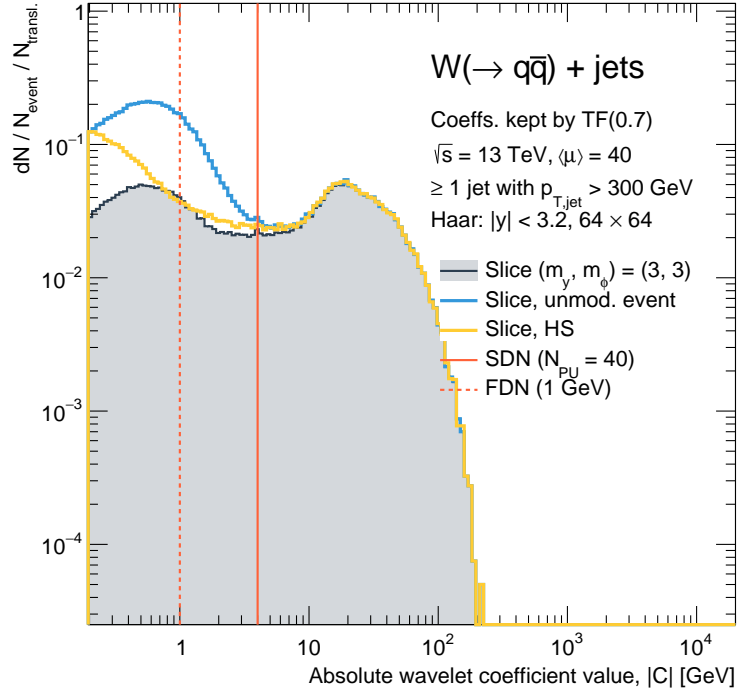

hard scatter. So, the TF method keeps all the high-energy coefficients while removing the softer ones in a smart, per-coefficient way.

Notice here, that the actual coefficient values, as plotted in Fig. 3.19, were never considered in the TF method, cf. Fig. 2.9. We never guarded coefficients with large energy content from being discarded, neither did we tell the method to predominantly get rid of the soft junk. We only relied on the corresponding track coefficients to guide the selection of coefficients.

This can be compared to the typical cuts of the FDN and SDN methods. These act by killing all coefficients to left of their respective cut values. However, as mentioned above, the hard scatter interaction alone might have some structure, encoded in the less-energetic coefficients, which it would be too harsh to discard completely. On the contrary, based on the idea of sparsity, cf. Sec. 2.2, one might reasonably argue that the structure of the hard scatter is faithfully captured in the high-energy coefficients alone, and thus that the softer contributions are indeed completely disposable, which to a large extent is true. Nevertheless, if it is, in fact, possible to keep some of the soft coefficients which we have reason to believe can be attributed to the hard scatter, this might be information worth keeping.

If we *were* to perform a simple cut, the question would then be where to separate the high- from the low-energy coefficients. As suggested by Eq. (2.45), this value might very well be dependent on the scales $m_{y,\phi}$ of the basis function. In Fig. 3.19 the SDN cut seems reasonable: it keeps coefficients in the range where the hard scatter and full event distributions agree, and discards coefficient with lower energies, where the two distributions diverge, indicating a considerable contribution from pile-up. However, this observation does not to hold at different scales, as shown in Fig. 3.20.

(a) At scales $(m_y, m_\phi) = (2, 2)$.
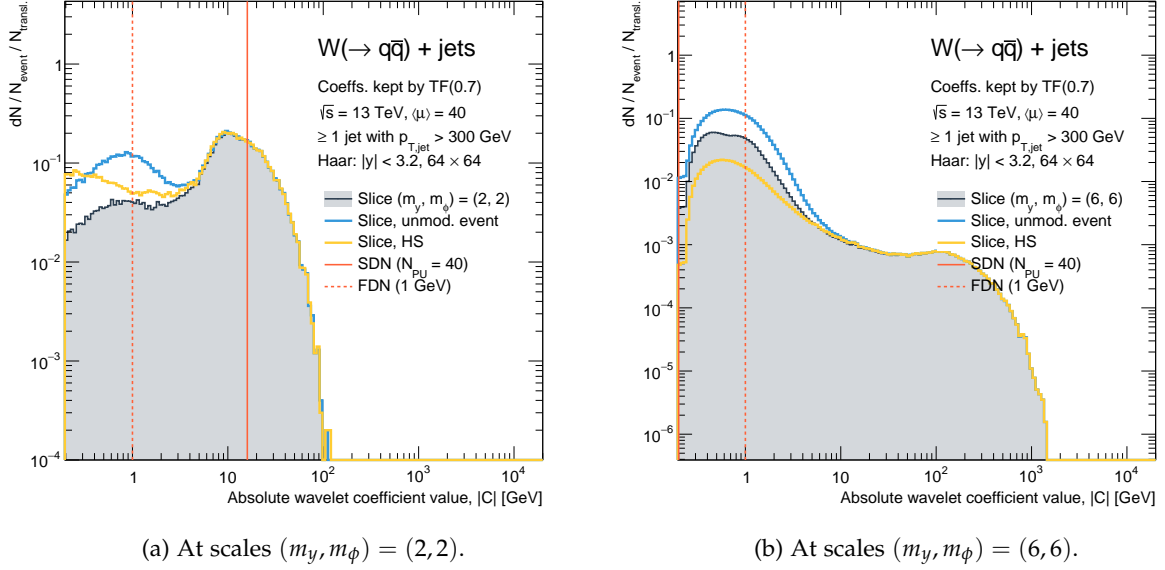
(b) At scales $(m_y, m_\phi) = (6, 6)$.

The two figures show the same distributions as Fig. 3.19, but at scales $(m_y, m_\phi) = (2, 2)$ and $(6, 6)$. Here we see, that the conclusions regarding the TF method holds: in the $|C|$ range, where the hard scatter and full event distributions agree, the TF method has no effect, but in the low-energy range where the two distributions disagree, it has the effect of removing ostensibly pile-up dominated coefficients. However, due to the scaling of the SDN cuts, cf. Eq.(2.45), this method is seen to have a markedly different effect. At low scales, Fig. 3.20(a), the cut seems too hard, effectively removing a large fraction of coefficients which are almost surely dominated by the hard scatter interaction. Conversely, at higher scales, Fig. 3.20(b), the SDN cut removes practically nothing. The fact that the SDN cuts have very differing effects at different scales may, however, be considered a "feature" rather than a "bug", cf. the remarks in Sec. 3.2. Nevertheless, across all three figures, the FDN cut, while a but low, seems far more able to demarcate hard scatter dominated coefficients from the more mixed ones. So, while neither of the two fixed-cut methods are immediately perfect, we it appears that a flat de-noising with a threshold of $5 - 10\,\text{GeV}$ holds definite potential. Similarly, the above discussion shows that auxiliary information, as utilised in the TF method, may allow us to discard individual coefficients on a well-motivated, per-event basis.

Having seen that the TF method performs well and in the way we hoped, we now want to gain a better understanding of the effect of the CL cut. Fig. 3.21 shows the distribution of output-to-input particle $p_\perp$ as a function of input $p_\perp$ when applying the TF(0.7) method, for particles from pile-up and from the hard scatter.

Below, we list a few striking characteristics of the plots:

- We have particles which have their momenta shifted below zero; these are the one which will be killed by the minimal, physically necessary threshold in the last step in Fig. 2.6. However, considering the harsh logarithmic scale, the number of particles in this region is small.

Figure 3.20: The distribution of coefficients kept by the TF(0.7) method (filled dark blue) in $W + \text{jets}$ events with the nominal setup cf. Sec. 3.1 at scales $(m_y, m_\phi) = (2, 2)$ and $(6, 6)$. Also drawn is the corresponding distribution in the full, unmodified event (full light blue) and the in the hard scatter event alone (full yellow) as well as typical SDN ($N_{\text{PU}} = 40$; full red) and FDN ($1\,\text{GeV}$; dashed red) cuts.

(a) Particles from pile-up.

(b) Particles from hard scatter.
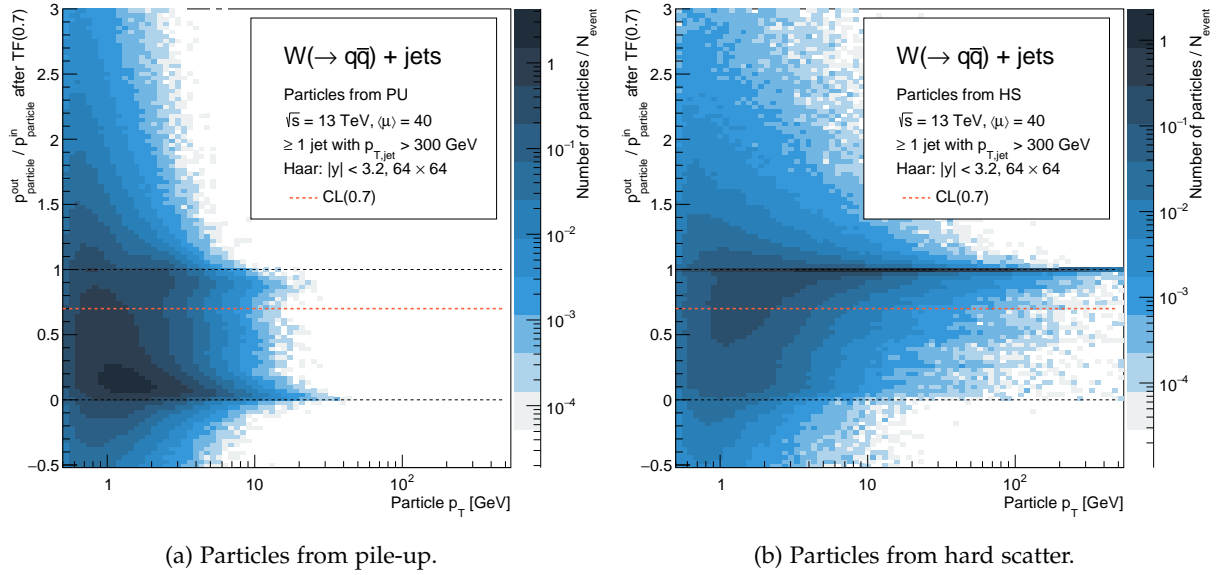
Figure 3.21: Output-to-input particle $p_\perp$ ratios, after application of the TF(0.7) method cf. Fig. 2.6, as a function of the input particle $p_\perp$, for pile-up and hard scatter particles, resp. Plots are made using the nominal setup, cf. Sec. 3.1. The CL(0.7) cut (dashed red) is shown.

- We similarly see that a small number of particles have their momentum increased (some considerably). This is a consequence of the general nature of wavelet analyses which, as discussed in Sec. 2.2, is acceptable as long as we make clear what we are dealing with after the wavelet analysis: not physical particles, but rather particle *proxies*.

- Pile-up particles are generally shifted towards the horizontal line at 0, while the hard scatter particles are shifted towards 1. This justifies the 'particle cleaning' "meta-method", since a flat cut as the one shown in Fig. 3.21 exactly has the effect of removing mainly pile-up particles.

- Finally, we see that the high-$p_\perp$ particles in the hard scatter event, thought to carry critical information about the event, are very sharply peaked at 1 and, as a result, are almost never killed by the cleaning. This explains why we are able to obtain such significant improvements in the jet mass resolution: we remove particle from pile-up (even ones width moderately high $p_\perp$), in a way that is fundamentally different from standard jet grooming techniques, while keeping almost all of the (high-$p_\perp$) hard scatter particles.

3.4   Results

In the previous sections we settled in a general structure for the wavelet analysis, optimised the two free parameters in simulation, reached the important conclusion that our optimised choice of wavelet analysis is robust to changes in boson flavour, jet $p_\perp$, and $\langle \mu \rangle$ just as it is well-justified based on its effect on wavelet coefficients and output particles. We now turn to the impact of this optimised wavelet analysis for physics results by examining individual event displays as well as the impact on our two chosen metrics: jet energy resolution and boson jet mass reconstruction, compared to standard methods. Finally we discuss various validation runs to assess whether the results are setup-dependent or truly credible.

■   Event displays

One of the most immediate ways to study the effect of our optimised wavelet analysis on hadron collision events is to do just that: look at the events themselves.

$\langle \mu \rangle = 40$

A representative $W + \text{jets}$ event with $p_{\perp,\text{jet}} > 300 \, \text{GeV}$ and $\langle \mu \rangle = 40$ is shown in Fig. 3.22.
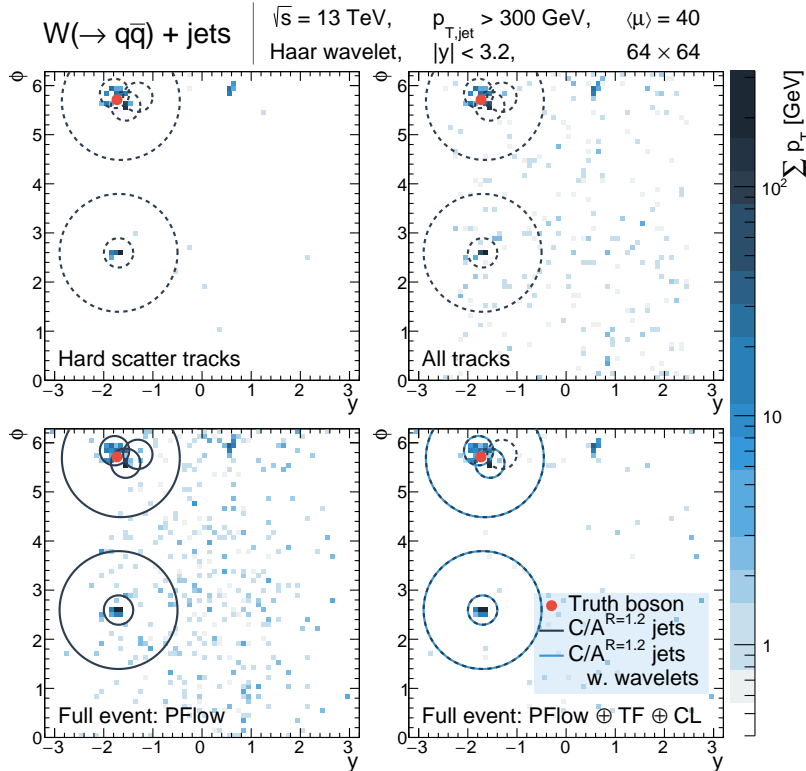

Figure 3.22: Display of $W + \text{jets}$ event with $p_{\perp,\text{jet}} > 300 \, \text{GeV}$ and $\langle \mu \rangle = 40$ and the nominal wavelet setup, cf. Sec. 3.1.

All four sub-figures are $y - \phi$ displays of the $p_\perp$ deposits in the detector, pixelised according to the nominal pixel definition (64×64 pixels out to $|y| < 3.2$).

The top left plot shows the distribution of tracks associated to the hard scatter vertex[14]. Also shown is the generator-level coordinates of the W boson as well as circular contours with radii 1.2 (0.3) showing the position and approximate size of the $C/A^{R=1.2}$ + BDRS-A jets (subjets) clustered in the full, unmodified event.

The top right plot shows the distribution of all tracks in the event. It is seen how pile-up introduces numerous soft tracks, roughly uniformly distributed in $y - \phi$ within the ID $|\eta|$ coverage, on top of the hard scatter ones.

The bottom left plot shows the distribution of particles flow-like objects (PFlow) for the full event, crudely corresponding to what the ATLAS calorimeter and tracking systems would be able to measure. At an average pile-up multiplicity[15] of $\langle \mu \rangle = 40$, we see the sizable soft, uniform noise obscuring the hard scatter event. Here the (sub-)jets contours are drawn in full dark blue. Since the uppermost jet is close in dR to the generator-level boson, this is what would be labeled a "W-jet".

Finally, the bottom right plot shows the PFlow object distribution *after* the application of the $TF(0.7) \oplus CL(0.7)$ analysis, which uses the information in the top left and right plots as auxiliary information. Jets clustered in the cleaned event are drawn in light blue. Since there is a clear overlap between the two sets of jets, i.e. it is clear which ones correspond to each other based on their proximity in dR, we can associate such pairs of similar jets and use it when studying the effect of wavelet-based subtraction on jet energy resolution. From this plot, the pile-up mitigating effect of the wavelet analysis is clearly visible. We see that the PFlow object deposits to the right of the display are mostly gone, while the high-$p_\perp$ deposits in the central parts of the jets are largely unmodified. However, we still see that the W-jet is clustered with one less subjet in the cleaned event, showing how wavelet methods have subtle yet important impacts on the kinematics of the reconstructed jets. The aim is then, that the jets reconstructed from the bottom right event display faithfully represent the particles which initiated them more so than the similar jets in the bottom left display.

The information shown in Fig. 3.22 is all something which is accessible in real data (apart from the truth-level vector boson), and the conclusions here, and in the following, should therefore also be applicable to real detector events, modulo inefficiencies, additional smearing, etc.

$\langle \mu \rangle = 200$

Fig. 3.23 shows the same hard scatter interaction overlaid with $\langle \mu \rangle = 200$ pile-up collisions.

Here we see the trouble caused by large amount of simultaneous collisions. Using the same $p_{\perp jet}$ threshold, four jets are now reconstructed in the unmodified event, as compared to the two jets in Fig. 3.22. But looking at the top left and right plots, we see the
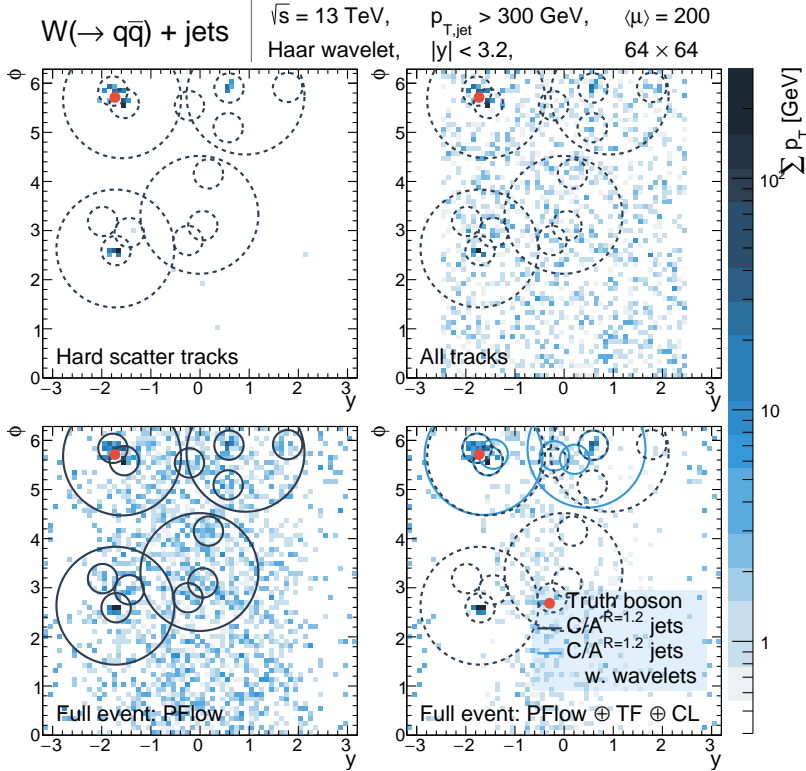
Figure 3.23: Display of W + jets event with $p_{\perp \text{jet}} > 300\,\text{GeV}$ and $\langle \mu \rangle = 200$ and the nominal wavelet setup, cf. Sec. 3.1.

promise of track-based subtraction: the large number of pile-up collisions result in a huge amount of additional tracks in the event, but the hard scatter tracks are the same. Therefore, the track-vertex matching has potential to remove at least some amount of the many PFlow objects from pile-up. This is what is shown in the lower right plot. Here we see the cleaning effect, particularly powerful in situations with this extreme an amount of pile-up. We see than two of the jets reconstructed in the unmodified event now have disappeared, leaving only the W- and one additional jet. We also see why this jet remains: due to the cluster of hard scatter track activity at $(y, \phi) \sim (1, 6)$, the TF method will tend to keep coefficients localised to this area of the detector. Therefore, less subtraction is performed here, and the jet retains sufficient energy to pass the $p_\perp$ requirement. This has little effect on the results for the W-jet itself. Another effect of this spurious cluster is the band of kept PFlow objects at $y \approx 0$ in the lower right plot: wavelet bases with $(m_y, m_\phi) = (0 - 1, 2 - 3)$ may pick up sufficiently large contribution from the hard scatter track cluster at $(y, \phi) \sim (1, 6)$, relative to the pile-up track, to pass the TF cut, thereby being more likely to retain particles within their support. Notice, however, how these effects could be avoided completely by discarding all primary vertex tracks sufficiently far away $(dR \approx 1.2)$ from the two largest energy deposits, e.g. by labeling as pile-up instead. This would likely kill the last non-resonant jet, and remove additional noise from within the W-jet, resulting in an even cleaner output event. However, such ideas are not pursued further in this context, as we will try to keep our methods as simple and transparent as possible.

Additional event display, including ones for higher-$p_\perp$ jets are

shown in App. A.2.

Having studied the impact of our chosen wavelet analysis on individual events, we now turn to its effect on our chosen metrics, cf. Sec. 3.1: jet energy resolution and boson jet mass reconstruction.

■ Jet energy resolution

We want to study the effect of the chosen wavelet analysis on our ability to reliably reconstruct the energy of jets, even in cases of large amount of pile-up. Our approach is described in Sec. 3.1: we cluster jets in the hard event, overlay it with some number of pile-up collisions and perform the jet clustering again; jets matched with $dR < 0.6$ are paired, and we then use Eq. (3.1), referred to as the 'jet energy resolution' (JER), to measure the residual difference in energy stemming from the pile-up collisions. Since the impact of pile-up removal on the JER is not specific to boson jets, but valuable also for QCD jets, for the reasons mentioned in Sec. 3.1, we use QCD dijet events in the following.

As we have mentioned earlier, wavelet-based methods do not exist in a vacuum, and we must take into account the several existing methods for pile-up removal and jet grooming when assessing their effect. For jet energy corrections, the most standard method is area subtraction, cf. Sec. 1.4. In the discussion of the JER metric, we will be using this method as a reference. In the present study, the median $p_\perp$ density has been computed based on a regular grid as in Eq. (1.60b), using the `fastjet::GridMedianBackgroundEstimator` class in FastJet [77] with rapidity range $|y| < 4.9$ and requested grid spacing $2\pi/10$.

Distributions

Distributions of the JER in QCD dijet events for $p_{\perp\text{jets}} > 300$ and $1000\,\text{GeV}$ with $\langle\mu\rangle = 40$ and 200, resp., are shown in Fig. 3.24.

Ideally we want a distribution which is centered around 0 (un-biased) and which is as narrow as possible (low RMS[16]). From Fig. 3.24(a) we see that, at $\langle\mu\rangle = 40$, the standard area subtraction yields improved results over the uncorrected jets, both reducing the distribution bias and width. The same is the case for $\langle\mu\rangle = 200$, albeit less pronounced due to the axis ranges. However, we clearly see the benefit of the wavelet analysis: the resulting distribution for $\langle\mu\rangle = 40$ has effectively removed any bias in the jet energy measurement, restored a gaussian shape, and narrowed the distribution width by $30 - 40\%$. This effect largely carries over in the case of $\langle\mu\rangle = 200$, however, with more skewed distribution.

Fig. 3.24(b) shows the same distribution, but with a $p_{\perp\text{jet}} > 1000\,\text{GeV}$ requirement. Even though the baseline-performance is much better, since the relative energy contribution from pile-up decreases and since the ATLAS calorimeter resolution improves at high $p_\perp$ cf. Sec. 1.3 and the discussion in Sec. 3.3, the conclusions are the same: area subtractions gives an improvement in both bias and RMS over uncor-
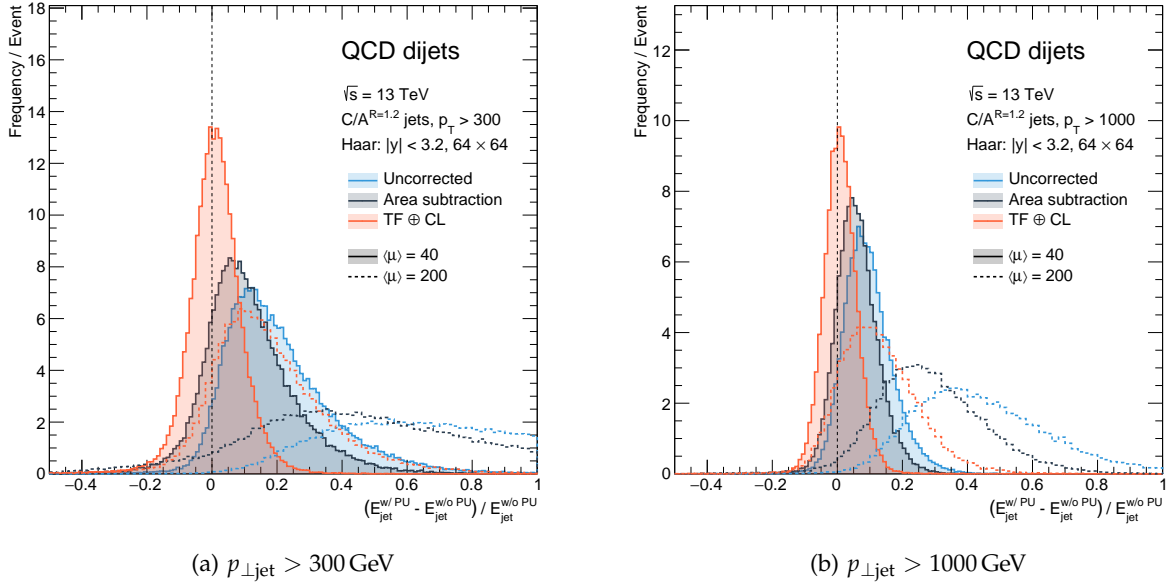
[16] Here, and in the following, 'RMS' actually denotes the standard deviation of the distribution, i.e.

$$\text{RMS} \equiv \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$$

with

$$\mu = \frac{1}{N}\sum_{i=1}^{N}x_i$$

This is done mainly to comply with ROOT nomenclature.

(a) $p_{\perp\text{jet}} > 300\,\text{GeV}$

(b) $p_{\perp\text{jet}} > 1000\,\text{GeV}$

rected jet, but is surpassed by the wavelet-based subtraction, both at moderate and extreme pile-up multiplicities.

   This indicates that our wavelet analyses can improve the measurements of jet energies both on average (bias) and on a per-event basis (RMS). This is promising, since the TF $\oplus$ CL parameters used presently were optimised for the boson jet mass metrics, the implication being that our choice of passes and parameter values is fairly robust across various measures, in addition to being performant.

Dependencies

In Fig. 3.24 we have seen JER distributions for four different $p_{\perp\text{jet}}/\langle\mu\rangle$ configurations. In order to study the behaviour of each method as functions of either, we plot the distribution bias (mean) and RMS as functions of $\langle\mu\rangle$ in Fig. 3.25.

   Both plots show a very simple, mostly linear structure: The bias grows linearly from zero at low $\langle\mu\rangle$ as expected from $\mathrm{d}\langle\rho\rangle/\mathrm{d}N_{\text{PU}} \approx 0.4\,\text{GeV}$, cf. Sec. 1.4, for all three $p_{\perp\text{jet}}$ cuts, with the wavelet-based subtraction giving the lowest slope in all cases. For $p_{\perp\text{jet}} > 1000\,\text{GeV}$, the wavelet analysis lowers the bias, compared to area subtraction, by figures ranging from $(\text{bias}_{\text{areasub.}} - \text{bias}_{\text{wavelet}})/\text{bias}_{\text{areasub.}} \approx 90\%$ at $\langle\mu\rangle = 20$ to $\approx 50\%$ at $\langle\mu\rangle = 200$. For lower $p_{\perp\text{jet}}$ thresholds the results are even stronger.

   However, the bias may be remedied by an average correction of the JER[17]. Therefore, it may be more interesting to study the per-event JER as quantified by the the RMS of the distribution. In Fig. 3.25(b), the RMS'es are seen to follow similar, linear evolutions with $\langle\mu\rangle$. Here we find no real improvements at $\langle\mu\rangle = 20$ for $p_{\perp\text{jet}} > 1000\,\text{GeV}$, but for $\langle\mu\rangle = 200$ wavelet-based subtraction lowers the RMS by $\approx 25\%$ compared to area subtraction. Similarly to the bias, we see further improvements at lower $p_{\perp\text{jet}}$ thresholds.

Figure 3.24: Jet energy resolution in QCD dijet events using the nominal setup, cf. Sec. 3.1. Distributions for jets after TF(0.7) $\oplus$ CL(0.7) subtraction (ref), after area subtraction (dark blue), and for uncorrected jets (light blue) are plotted at $\langle\mu\rangle = 40$ (filled) and 200 (dashed).

[17] This is the principle behind the ATLAS "offset correction" [78].

(a) Bias

(b) RMS
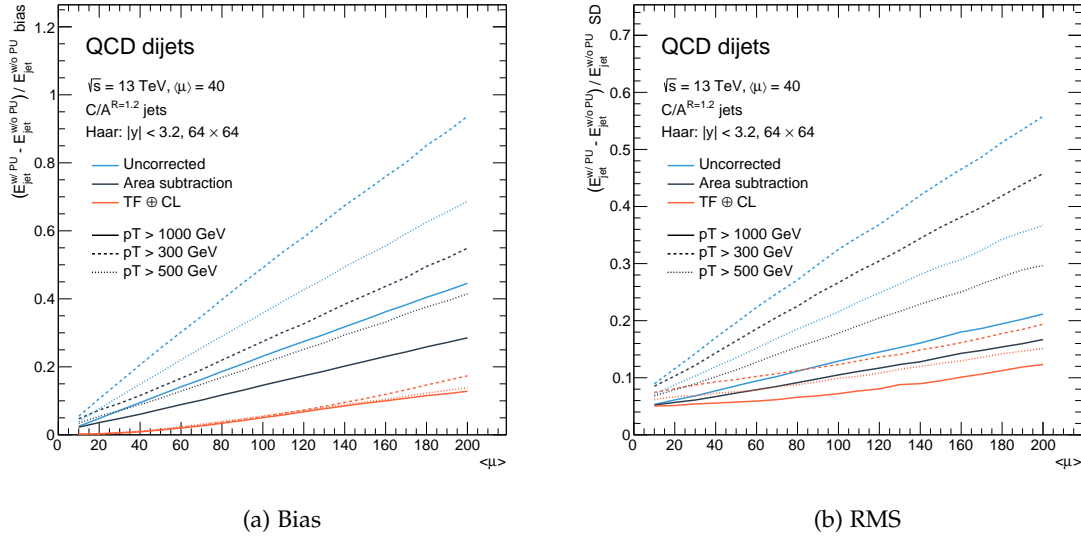
Figure 3.25: Jet energy resolution bias and RMS as functions of $\langle \mu \rangle$ in QCD dijet events, otherwise using the nominal setup, cf. Sec. 3.1. Distributions for jets after $\mathrm{TF}(0.7) \oplus \mathrm{CL}(0.7)$ subtraction (red), after area subtraction (dark blue), and for uncorrected jets (light blue) are plotted for three different $p_{\perp \mathrm{jet}}$ cuts.

(a) Bias

(b) RMS
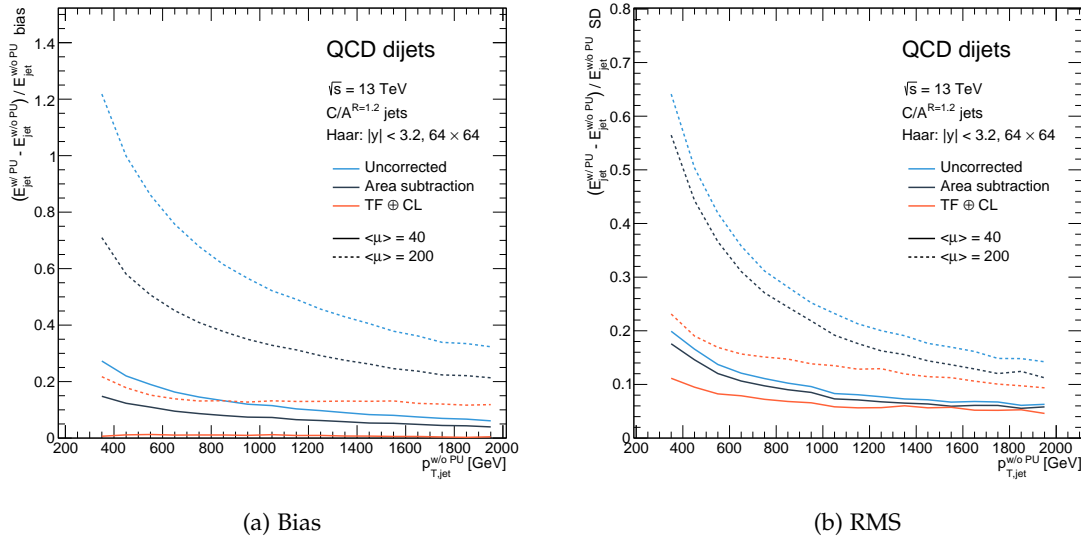
Figure 3.26: Jet energy resolution bias and RMS as functions of $p_{\perp \mathrm{jet}}^{\mathrm{w/o\,PU}}$ in QCD dijet events, otherwise using the nominal setup, cf. Sec. 3.1. Distributions for jets after $\mathrm{TF}(0.7) \oplus \mathrm{CL}(0.7)$ subtraction (red), after area subtraction (dark blue), and for uncorrected jets (light blue) are plotted for two different average pile-up multiplicities.

In Fig. 3.26 we plot the distribution bias and RMS as functions of the transverse momentum of the pile-up-free jet, $p_{\perp \mathrm{jet}}^{\mathrm{w/o\,PU}}$.

This plots shows the behaviour of the JER distributions with the transverse energy scale of the event. We recognise the same behaviour as noted above: at high jet energies, the natural calorimeter resolution improves the measurement and the energy contribution from pile-up becomes comparatively smaller, reducing the potential for improvement from pile-up removal methods. Nevertheless, we see

that wavelet-based subtraction yields consistent improvement across all studied jet transverse momenta, both at moderate ($\langle\mu\rangle = 40$) and more extreme ($\langle\mu\rangle = 200$) pile-up multiplicities. Finally, we notice the regularity of the graphs both in Fig. 3.25 and Fig. 3.26: the track-guided subtraction has an almost identical functional behaviour as the standard methods, as functions of both pile-up activity and transverse energy, despite its arguably more convoluted nature, cf. Fig. 2.6.

The plots in Figs. 3.25 and 3.26 represent orthogonal slices in a more general 2D picture. Since we have noted that the width of the JER distribution is probably the most relevant quantity to study, Fig. 3.27 shows the ratio of the number of wavelet-corrected jets to jets corrected by area subtraction, within a symmetric ± 0.15 JER window around the respective distribution maxima, as functions of $p_{\perp\text{jet}}^{\text{w/o PU}}$ and $\langle\mu\rangle$. We choose this metrics since it is simple (requiring only a counting within a fixed-width interval), captures information about the width of the distribution regardless of the bias, and, due to low statistics in the high-$p_\perp$ region, it avoids un-physical tail biases otherwise encountered with the RMS.
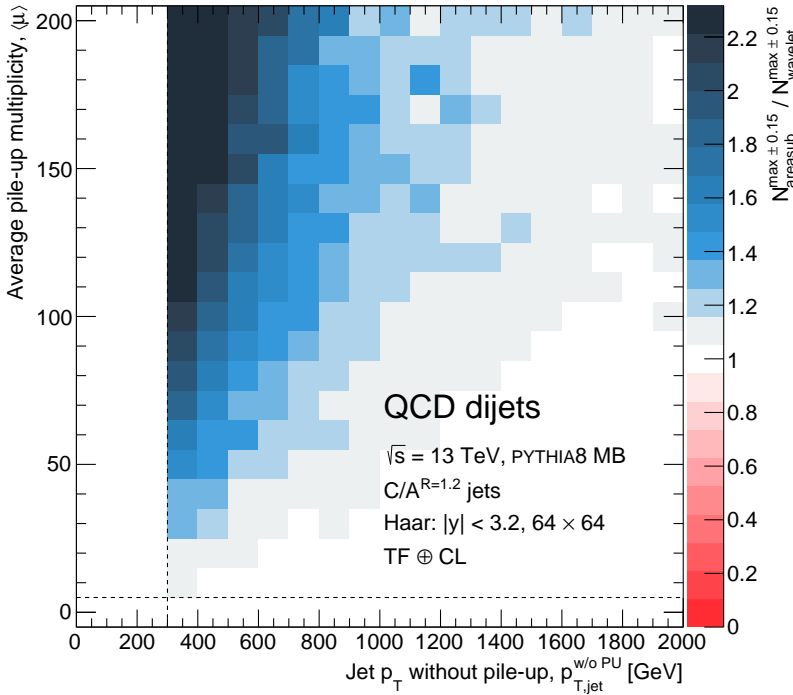
Figure 3.27: Improvement in the number of jets within a fixed interval of ± 0.15 around the distribution maximum for jets corrected by wavelet-based (TF $\oplus$ CL) subtraction relative to four-vector area-subtracted jets in QCD dijet events.

Here we again see the regular behaviour of the improvements obtainable by using wavelet-based subtraction: in the low-$p_\perp$/high-$\langle\mu\rangle$ limit, sizable improvement may be achieved and in the high-$p_\perp$/low-$\langle\mu\rangle$ the subtraction is on par with existing methods. Generally we find that, across all $p_{\perp\text{jet}}/\langle\mu\rangle$ configurations, wavelet-based subtraction performs as-well or better than standard techniques. The fact that the effect of the wavelet-based subtraction approaches that of area subtraction is not trivial, since the principles of operation for the two methods are markedly different, as are their underlying assumptions.

However, both methods are, by construction, safe[18] in the limit of $\langle \mu \rangle \to 0$. Similarly, the relative energy contribution from pile-up vanishes in the high-$p_{\perp \text{jet}}$-limit, which is also true for both methods. Therefore, the smooth convergence in the high-$p_{\perp}$/low-$\langle \mu \rangle$ limit may be less of a (welcome) surprise.

It should be noted that the area subtraction used here is not the full method as used in official ATLAS reconstruction. There, smart residual corrections are also applied [55]. However, the area four-vector subtraction, based on the median $p_{\perp}$ density, is intended to remove the diffuse contribution from pile-up in a simple, per-event manner. Exactly the same is true for the wavelet-based subtraction. Therefore, although some more involved corrections have been omitted, which could be applied to either of the two method, we believe that the above constitutes a reasonably fair comparison between the different techniques.

Notice, that the above results hinge on the exception for the $(m_y, m_\phi) = (0, 0)$ basis function in Fig. 2.9. If this had not been made, we would frequently be discarding this coefficient, having an erroneous per-event jet energy normalisation, and correspondingly the improvements found above would vanish.

It is worth noting that the wavelet analysis used here was optimised for boson jet mass reconstruction. The results shown in Figs. 3.25 and 3.26 for both the distribution bias and width are therefore not the result of a fine-tuning, but derived effects of an evidently quite robust method.

■ Boosted boson jet mass

We now turn to our other metric: the reconstructed mass of boson jets. We cluster jets in the original event as well as in the wavelet-modified event, cf. Fig. 2.6, and those falling within $dR < 0.6$ of the truth boson are labeled 'boson jets', as mentioned above. This is done using the nominal wavelet and jet setup, sampling signal (boson) jets from $W +$ jets events and background (non-boson) jets from QCD dijet events. We use standard BDRS-A jet grooming as our basis of comparison, such that any in difference performance will be the result of the wavelet-based cleaning.

Distributions

We plot the jet mass, M_{jet}, distributions in Fig. 3.28 for jets with $p_{\perp} > 300\,\text{GeV}$, where we also show the bare (i.e. ungroomed) jet mass distributions.

It is immediately clear from Fig. 3.28(a) that the BDRS-A filtering does an impressive job: it shifts the mean from $\sim 200\,\text{GeV}$ down to a value of $\sim 90\,\text{GeV}$, already much more reflective of the boson pole mass (80.385 GeV [2]; indicated by the dashed black line). Additionally, where the bare QCD background was previously practically indistinguishable from the signal distribution, the BDRS-A filtering also reduces the signal distribution width substantially, and more so

(a) $\langle \mu \rangle = 40$.

(b) $\langle \mu \rangle = 200$
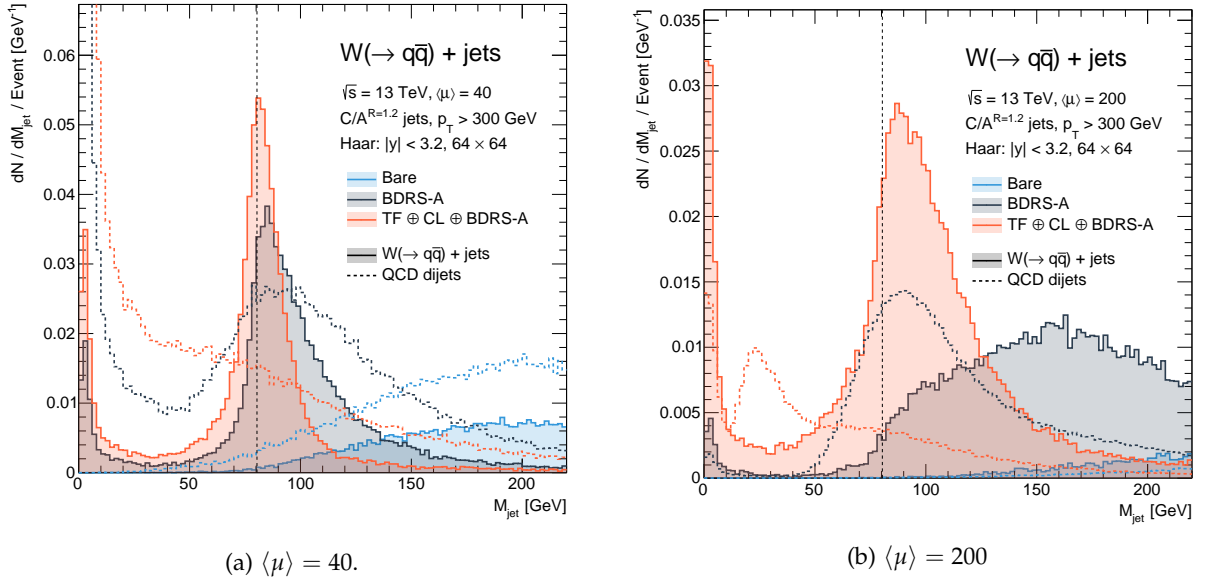
than for the background distribution, to a level where a cut window seems feasible. Finally, we see the steep rise as M_{jet} goes to zero, corresponding to jets which are "killed" by the grooming procedure, showing why the $M_{jet} > 20\,\mathrm{GeV}$ cut in the optimisation procedure, cf. Sec. 3.3, was warranted.

But Fig. 3.28(a) also shows that the wavelet analysis does *its* job: where the moderate pile-up multiplicity $\langle \mu \rangle = 40$ already biases the BDRS-A groomed distribution slightly towards larger M_{jet} values and widens its shape, the track-guided subtraction removes (and scales) the soft parts of the event to an extent, where the standard grooming can perform better, resulting in virtually no bias as well as a reduction in the distribution width (FWHM) of $\approx 25\%$. What is more, the QCD background is changed in two important ways: first, its shape becomes far more regular, having no bump behaviour[19] in the W pole mass region, and second, the number of QCD jets in the region around the boson jet mass peak (e.g. $M_{jet} \in [60, 100]\,\mathrm{GeV}$) is reduced by roughly one third, giving further separating power against continuum jets.

Here it is also worth pointing out, that the optimisation in Sec. 3.3 was based solely on the distribution width (modulo the $M_{jet} > 20\,\mathrm{GeV}$ requirement). The fact that the wavelet analysis has the additional effect of removing (or reducing) the bias is unintended but welcome. This serves to show that the results of the wavelet-based subtraction are not strange chance effects of a careful tuning under very particular circumstances (see $p_\perp / \langle \mu \rangle$ dependence below). Instead, the subtraction seems to work in the way it was intended: reliably removing pile-up, such that mainly the hard scatter particles remain, thereby *naturally* reducing any bias introduced by the pile-up. Conversely, the subtraction doesn't undershoot either, introducing a negative bias, confirming that it removes essentially *only* pile-up.

These conclusions all carry over in the case of $\langle \mu \rangle = 200$, as shown in Fig. 3.28(b). However, we want to repeat the remarks of Sec. 3.3,

Figure 3.28: Jet mass distributions for boosted W boson jets (filled) compared with QCD jets (dashed) for the nominal setup, cf. Sec. 3.1. Distributions for the bare jets (light blue), groomed jets (dark blue), and groomed jets clustered in the wavelet-cleaned event (red) are shown. The dashed black line marks the W boson pole mass.

[19] The bump, and its position, is an artifact of the $p_{\perp jet}$ cut: if the $p_\perp$ cut is increased, the bump is shifted towards larger M_{jet} values; if the cut is decreased, the shift is in the other direction, see e.g. Fig. 4.8(a). This means that, in any analysis using an M_{jet} cut to select hadronically decaying boson jets, the prior $p_{\perp jet}$ cut ought to be chosen with care, so as to not introduce an artificial bump-structure in the background distribution.

urging moderation in our optimism: the standard grooming methods, and in particular BDRS-A, we're never intended to perform well under conditions of such extreme pile-up activity, as is clear from the figure. However, the plot also shows that a simple $TF \oplus CL$ analysis is able to perform sufficient pile-up removal that the standard grooming methods can perform acceptably. Therefore we remain cautiously optimistic even for such extreme run conditions.

However, as remarked in Sec. 3.3 and the above discussion of the JER results, the improvement from track-guided wavelet subtraction disappears in cases with high-$p_\perp$/low-$\langle\mu\rangle$. This is shown in Fig. 3.29.

Figure 3.29: Jet mass distributions for boosted W boson jets (filled) compared with QCD jets (dashed) requiring $p_{\perp\mathrm{jet}} > 300\,\mathrm{GeV}$ and otherwise using the nominal setup, cf. Sec. 3.1. Distributions for the bare jets (light blue), groomed jets (dark blue), and groomed jets clustered in the wavelet-cleaned event (red) are shown. The dashed black line marks the W boson pole mass.
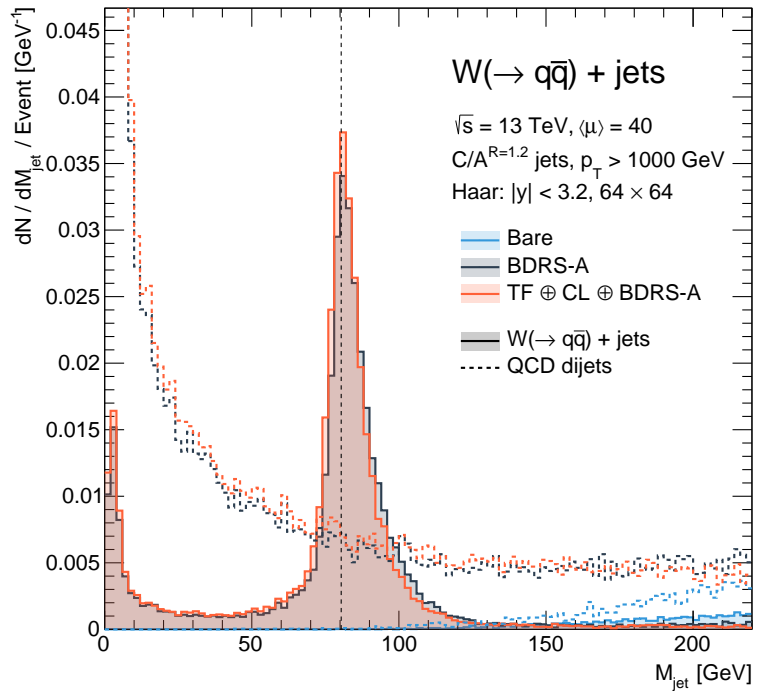
Here we see that the wavelet analysis is not able to offer any further improvement over standard BDRS-A grooming, for neither the signal nor the background distribution. As was also the case with the jet energy resolution, we see that the large transverse momenta of the jets yields smaller relative impact of the pile-up contribution as well as a better calorimetric resolution, resulting in a better starting point for the jet grooming. Furthermore, the higher-$p_\perp$ jet constituents also serve to make the grooming methods more effective against the increasingly soft pile-up contamination, giving an overall better jet mass resolution–even in the absence of any additional cleaning. But an important observation is also, as noted above, that performance doesn't deteriorate either: despite the TF method relying solely on track information to discard or keep individual coefficients, with no regards for their energy/information content, the hard part of the event is left essentially unchanged, resulting in a jet mass distribution which is no worse than for standard methods alone. This confirms the robustness of the chosen wavelet analysis structure.

Dependencies

Similarly to the study of the JER results above, we want to quantify the improvement offered by the wavelet analysis and study its dependence on the two major parameters: $p_{\perp\text{jet}}$ and $\langle\mu\rangle$. To do this, we compute the $S/\sqrt{B}$ for the TF $\oplus$ CL $\oplus$ BDRS-A distributions in the fixed range $M_{\text{jet}} \in [60, 100]$ GeV and takes its ratio to $S/\sqrt{B}$ value for the BDRS-A–only distribution on the same range. This figure of merit is similar to that used in the JER discussion above, involving only jet counts within a fixed-width interval. However, we choose to center the interval on (roughly) the W pole mass, instead of on the maxima of the respective boson jet mass distributions, since one usually want a simple, physically motivated cut on the jet mass without resorting to "offset corrections", which are difficult to compute due to the rarity of boson jets. Also, as noted in Sec. 3.1, this measure also takes into account any loss of efficiency caused by the wavelet cleaning. Fig. 3.30 shows the dependence of this measure on the jet $p_{\perp}$ for five different slices in $\langle\mu\rangle$.

Figure 3.30: Ratio of $S/\sqrt{B}$ for M_{jet} distributions after wavelet analysis and jet grooming to those after only jet grooming, for boson vs. QCD jets, computed in the fixed $M_{\text{jet}} \in [60, 100]$ GeV interval, as function of the jet $p_{\perp}$ for five average pile-up multiplicities $\langle\mu\rangle$. The shift in the left $p_{\perp\text{jet}}$ limit for each contour is introduced to avoid edge-effects at increasing levels of pile-up, and has magnitude $\sim 0.4\,\text{GeV} \times \langle\mu\rangle \times \pi\,1.2^2$, cf. Sec. 1.4.

Here we see a behaviour very similar to that found in Fig. 3.26: wavelet-based subtraction generally has a decreasing effect at larger jet $p_{\perp}$, but the absolute improvement increases with $\langle\mu\rangle$, also shifting upwards in $p_{\perp}$ the point where the subtraction becomes ineffective. This point is seen to change roughly as $\sim 25\,\text{GeV} \times \langle\mu\rangle$. Since we find little improvement for $p_{\perp\text{jet}} \gtrsim 600\,\text{GeV}$ for LHC Run 1 pile-up conditions, cf. Sec. 2.2, we see that the real promise of wavelet-based subtraction lies in Run 2 and beyond.

Fig. 3.31 shows the dependence of the relative $S/\sqrt{B}$ measure on both $p_{\perp\text{jet}}$ and $\langle\mu\rangle$.
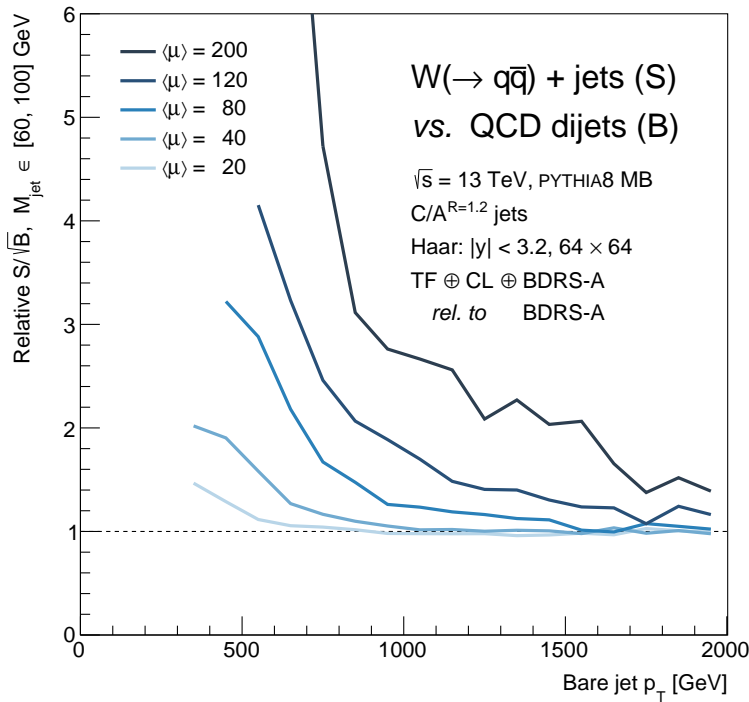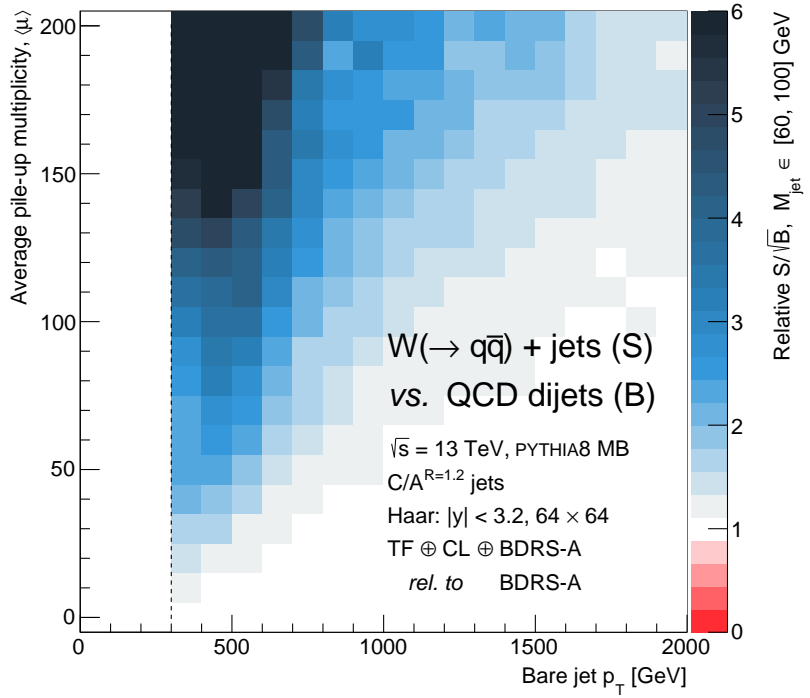
Figure 3.31: Ratio of $S/\sqrt{B}$ for M_{jet} distributions after wavelet analysis and jet grooming to those after only jet grooming, for boson vs. QCD jets, computed in the fixed $M_{\text{jet}} \in [60, 100]\,\text{GeV}$ interval, as function of the jet $p_\perp$ and the average pile-up multiplicity $\langle \mu \rangle$.

[20] Both plots are truncated at improvements of a factor of six.

We see that the improvements follows the same general behaviour as those for the JER, cf. Fig. 3.27: we see an improvement in performance for low-$p_\perp$/high-$\langle \mu \rangle$, which deteriorates smoothly going to high-$p_\perp$/low-$\langle \mu \rangle$, while never resulting in a performance worse than standard methods alone. One difference seen here[20], as well as in Fig. 3.30, is that the *magnitude* of the improvements is much larger here than for JER. However, this is merely an expression of the fact that the present measure of improvement takes the distribution bias, as well as its width, into account (using an fixed interval), while the JER measure only accounts for improvements in width (using a fixed-width interval with variable position). The choice of measure is somewhat arbitrary. We have tried to chose two, which reflect the needs in physics analyses. However, the two (or other) measures can be used interchangeably for either of the two metrics, and the plots in Figs. 3.27 and 3.31 will change accordingly.

Initially, one might wonder why, even at moderately low $\langle \mu \rangle$, the wavelet-based subtraction is able to improve the performance of standard methods: jet grooming works to remove soft constituents of the jet, and does so well, so why should/does the wavelet analysis make a difference? The key point is, that the two methods work in very different ways. Jet grooming examines individual jet constituents or protojets to see if they are sufficiently soft or collinear to be discarded. This is a highly localised, binary, iterative procedure intended to remove the soft parts (often thought of only as UE) of an already clustered jets. Wavelet-based subtraction, however, works at the level of the entire detector, without reliance on specific clustering algorithms and on he kinematics of individual particles. Our chosen wavelet setup seeks to remove the soft contributions not coming form

the primary vertex in a function space which is fundamental different in character from the standard $y - \phi$ space. The result, as exemplified by Fig. 3.31, is that the two methods are not completely correlated and that the combination of the two gives results which it would not be possible to obtain with either methods on its own.

■ Validations

Having studied the performance of our chosen wavelet analysis as it manifests itself in the two chosen methods, jet energy resolution and boson jet mass measurement, we also need to confirm that these results are not specific to our chosen setup, cf. Sec. 3.1. To do this, we perform a similar study as for the boson jet mass above, using the nominal setup with $W + $ jets events with $\langle \mu \rangle = 40$ and requiring $p_{\perp\mathrm{jet}} > 300\,\mathrm{GeV}$, where we change one parameter in the setup to see how it affects the performance. This is shown in Tab. 3.1.

Setup	$S/\sqrt{B}$ wrt. nom.	$S/\sqrt{B}$ wrt. no wavelet
Nominal	1.000	1.60 (2)
Herwig++	0.965 (15)	1.68 (3)
32 × 32	0.840 (13)	1.36 (2)
128 × 128	1.006 (15)	1.64 (3)
256 × 256	0.929 (14)	1.49 (2)
Daubechies 4	0.966 (15)	1.55 (2)
Daubechies 8	0.937 (14)	1.52 (2)
C/A$^{R=1.2}$ + BDRS	1.104 (18)	1.19 (2)
anti-$k_\perp^{R=1.0}$ + trim.	0.960 (14)	1.289 (19)
C/A$^{R=0.8}$ + prun.	1.009 (16)	1.22 (2)

Table 3.1: Table showing relative $S/\sqrt{B}$ values for jet mass reconstruction with different validation setups. Both values relative to the nominal distribution and to distributions without pile-up subtraction are shown. Signal (S; W-jets in $W +$ jets events) and background (B; QCD jets in QCD dijet events) counts are computed for jets, with $p_{\perp\mathrm{jet}} > 300\,\mathrm{GeV}$ clustered in the $\mathrm{TF}(0.7) \oplus \mathrm{CL}(0.7)$ output and in the unmodified event, in the M_{jet} interval [60, 100] GeV, with $\langle \mu \rangle = 40$.

The first column shows the $S/\sqrt{B}$ for wavelet-cleaned jets relative to that in the nominal setup. The second column shows the $S/\sqrt{B}$ for wavelet-cleaned jets relative to that for un-cleaned (yet still groomed) jets[21]. The validation runs cover the minbias generator, the pixel resolution, the wavelet dictionary, and the jet clustering and grooming scheme. Numbers in parentheses show only the combined statistical uncertainty on the $S/\sqrt{B}$ ratio, found through standard error propagation[22], using

$$\sigma_{S/\sqrt{B}} = \sqrt{\left(\frac{1}{\sqrt{B}}\right)^2 \sqrt{S}^2 + \left(\frac{S}{2B^{3/2}}\right)^2 \sqrt{B}^2} = \sqrt{\frac{S}{B} + \frac{S^2}{4B^2}} \quad (3.4)$$

for each $S/\sqrt{B}$ value. Nevertheless, the goal is not a detailed error analysis, but rather just rough estimates of the consistency of our results.

Since the alternative minbias generator, here Herwig++, represents an alternative[23] to simulate soft QCD events, in particular the hadronisation cf. Sec. 1.2, we would expect similar performance for

[21] The jet mass reconstruction improvements over standard techniques shown in Tab. 3.1 for the nominal setup is somewhat lower than what one would seem to expect from Fig. 3.30. However, the events used in Tab. 3.1 has only a minimum $p_{\perp\mathrm{jet}}$, whereas each point in Fig. 3.30 corresponds to a $p_{\perp\mathrm{jet}}$ interval: the first point for the $\langle \mu \rangle = 40$ contour covers the interval [300, 400] GeV, and since this figure also has shown that improvements from wavelet analyses deteriorate with higher $p_{\perp\mathrm{jet}}$, this difference is to be expected.

[22] Assuming uncorrelated, Poissonian errors.

[23] Notice that the Herwig tune used in the present study, UE-E5, has shown to undershoot quite enormously when compared to 13 TeV data [79], and therefore any discrepancy found here may be, at least in part, attributed to this effect.

this validation setup and the nominal. Any discrepancy way be thought of as a systematic uncertainty associated with the wavelet-based pile-up removal. The different M_{jet} distributions for the two minbias generators are shown in Fig. 3.32.

Figure 3.32: Jet mass distributions for the nominal setup, cf. Sec. 3.1, using PYTHIA8 (red) to generate minimum bias events, compared to using HERWIG++ (green). The plots compares the performance of TF ⊕ CL ⊕ BDRS-A (filled) to that for BDRS-A alone (line) for *W* (full) and QCD (dashed) jets.
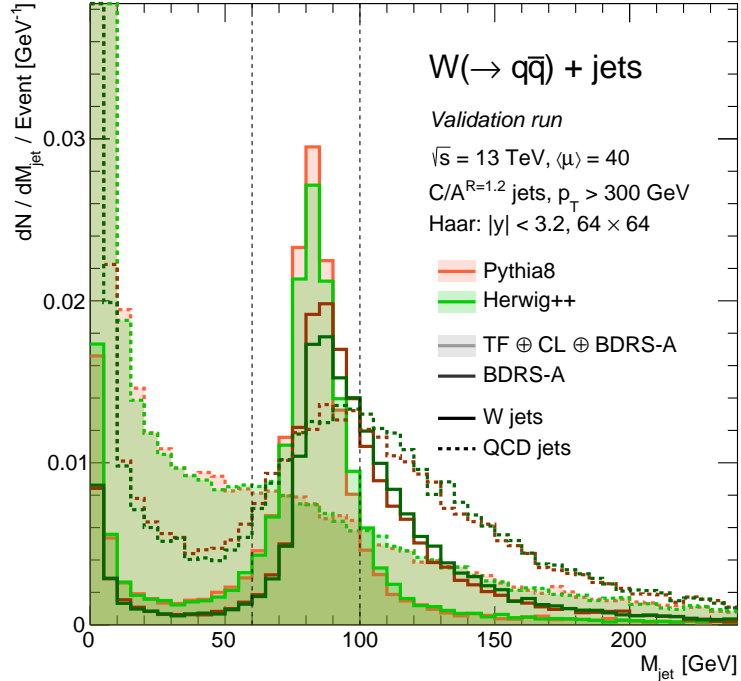
We see that, when comparing to the nominal setup, we find a slight deviation from unity, $0.965\,(15)$, which is not consistent with statistical errors alone, signaling that our chosen wavelet analysis fares slightly better in PYTHIA8 minbias. This is also the impression based on the two shaded *W* jets distributions in Fig. 3.32. Based on this result one might attribute a 3.5% systematic error, from the minbias modeling, to all wavelet-based cleaning results. We also see that the improvements in $S/\sqrt{B}$ for HERWIG++ and PYTHIA8, from wavelet-based cleaning wrt. grooming alone, are roughly consistent at the 2σ-level.

The remaining validation setups should not necessarily yield results which are compatible with the nominal setup, since each of these represent a different analysis approach: the pixel resolution, wavelet dictionary, and jet clustering and grooming scheme may be chosen freely and, although we don't expect the results to change too dramatically as a result, they will very likely differ.

The 32×32 setup shows a significantly reduced $S/\sqrt{B}$ relative to the nominal setup, but still gives a sizable improvement over standard grooming. This means that, although the coarser resolution gives better results than grooming alone, there is high-frequent information which is not fully utilised. This converse is not true for the higher-resolution setups. Here we see no improvement (or actual deterioration) relative to the 64×64 resolution. The seeming

decrease in performance for the 256×256 setup must be attributed to the highest-scale basis functions. At such high granularity, the pixel array will be so sparsely filled, that particularly basis functions with $m_y \ll m_\phi$ or $m_y \gg m_\phi$ are susceptible to chance configurations where individual pile-up tracks "accidentally" align with neutral hard scatter particles, effectively killing the latter. If this is the case, the loss of performance wrt. the 128×128 resolution might vanish at larger $\langle \mu \rangle$. However, this remains to be confirmed. It is an indication that higher resolution is not necessary beneficial, but should be chosen in accordance with the specific run conditions and analysis needs.

The use of different and more complex wavelet dictionaries actually seems to yield slightly smaller improvements. This might be somewhat surprising giving the fact that the Haar basis functions in Fig. 2.3 do not immediately seem optimal for describing jets. Similarly, it may be argued that the seemingly more pronounced two-prong structure of the Daubechies wavelets, see e.g. Fig. 2.1(b), ought to be better suited to capture the features of the two-subjet structure of hadronic vector boson decay. However, the difference in "prong-structure" between the Haar and Daubechies dictionaries in two dimensions is vanishing, and the results appear to favor the simpler alternative.

Finally, for the validation jet clustering and grooming schemes, we see that, compared to the nominal setup, no striking differences are found: $C/A^{R=1.2}$ + BDRS performs slightly better and anti-$k_\perp^{R=1.0}$ + trimming seems to perform slightly worse. However, all three validation setups fare markedly better for jets in the unmodified event, indicating that BDRS-A filtering is a sub-optimal grooming scheme under the studied circumstances. Nevertheless, we see that, regardless of clustering algorithm, distance parameter, and grooming technique, wavelet-based subtraction offers an event- and jet-cleaning which is not completely correlated with standard methods.

Based on this discussion, it seems that the wavelet and jet setup chosen presently (64×64 resolution, Haar dictionary, and $C/A^{R=1.2}$ + BDRS-A, cf. Sec. 3.1) seems sensible and, furthermore, we have gained confidence that the above results are not specific to this setup.

IV   Analysis: ATLAS simulation

4.1 Wavelets in ATLAS MC

In the previous sections we have studied the behaviour and performance of wavelet methods, in particular track-based ones, in toy simulation. However, if we are to employ wavelet analyses in searches for new physics, we must make sure that the results found so far are not due to simplistic modeling. In this section, we present the processes and samples used for studying wavelet methods in official ATLAS Monte Carlo (MC), as well as the complications which arise when we are faced with the complexity of actual (albeit simulated) detectors. In order to confirm that the basis for our wavelet analysis is sound, also in ATLAS MC, we study various wavelet coefficient distributions, and compare these to toy simulation. Finally, we apply our wavelet analysis to ATLAS MC events, and study its impact on the reconstruction of boson jet masses.

■ Processes and samples

We want to study processes resembling the ones in toy simulation, cf. Sec. 3.1, as much as possible. However, we are limited by the selection provided by the official ATLAS Run 2 simulation campaigns[1]. On one hand, we want samples containing hadronically decaying W bosons, and on the other we want samples containing QCD jets with an initiating-parton composition which reflects what is expected in actual data.

Since this study is motivated by searches for new physics in diboson final state, see Sec. 1.1, we use hard interaction samples containing an exotic resonance—specifically, a neutral heavy vector triplet, HVT—decaying semi-leptonically through two W's as HVT $\rightarrow WW \rightarrow \ell\nu q\bar{q}$. These processes contain a boosted W jet, and have the additional benefit of allowing us to relate the results of this section directly to the diboson search presented in Sec. 4.2. The resonances are produced at six different mass points, in $500\,\mathrm{GeV}$ steps from $M_{\mathrm{HVT}} = 500\,\mathrm{GeV}$, and we will interchangeably refer to the samples as e.g. HVT(1.0 TeV) or $WW_{(1.0\,\mathrm{TeV})}$, for a resonance mass of 1 TeV, depending on whether the emphasis is on the resonance search or on the W boson jet itself. These samples use the MSTW 2008 LO PDF set [80] with the AU2 underlying event tune [74] and MADGRAPH [81] matrix element with parton showering and hadronisation performed with PYTHIA8 [70]. For the background jets, we use the JZxW QCD dijet samples with x $= 2 - 7$, generated with HERWIG++ [24] using the CTEQ6L1 PDF set [82]. These are segmented in parton $\hat{p}_{\perp}$ [83] so as to allow for sufficient statistics even in the very boosted regions. All samples used are produced in the mc14 simulation project at $\sqrt{s} = 13\,\mathrm{TeV}$, processed in the xAOD format using the ∗_r5787_r5853 reco. tag. The simulation of the ATLAS detector is performed in GEANT4 [84], which simulates the passage of final state particles through the material of different detector components, resulting in a performance which is much closer to real data than our simple toy smearing, cf. Sec. 3.1. Each event is then processed by the full ATLAS reconstruction setup, as would real data. Basic information regarding the ATLAS dataset ID, kinematic region, the cross section times branching ratio, generator filter

[1] Presently we have only used samples from the mc14 campaign ('DC14'), since these were the ones available at the start of the project period.

efficiency, and number of generated events, for each sample used in the present study of our wavelet analysis' performance, is listed in Tab. 4.1.

Name	Dataset ID	M_{HVT} [TeV] or $\hat{p}_\perp$ range [GeV]	$\sigma \times \mathrm{BR}$ [fb]	$\varepsilon_{\mathrm{filter}}$	$N_{\mathrm{gen.}}$
HVT(0.5 TeV)	203481	0.5	1765.	7.5616×10^{-1}	30000
HVT(1.0 TeV)	203482	1.0	98.95	7.9607×10^{-1}	30000
HVT(1.5 TeV)	203483	1.5	16.11	8.1147×10^{-1}	29500
HVT(2.0 TeV)	203484	2.0	3.810	8.1310×10^{-1}	30000
HVT(2.5 TeV)	203485	2.5	1.076	8.1773×10^{-1}	30000
HVT(3.0 TeV)	203486	3.0	0.3344	8.1999×10^{-1}	30000
JZ2W	187032	[80, 200]	4.3883×10^{10}	3.2619×10^{-3}	500000
JZ3W	187032	[200, 500]	1.0089×10^{9}	1.2276×10^{-3}	499500
JZ4W	187034	[500, 1000]	1.5338×10^{7}	9.9168×10^{-4}	500000
JZ5W	187035	[1000, 1500]	4.3570×10^{5}	1.4351×10^{-3}	500000
JZ6W	187036	[1500, 2000]	4.2693×10^{4}	9.8697×10^{-4}	499500
JZ7W	187037	> 2000	7002.5	5.2787×10^{-3}	499500

Table 4.1: Brief information about the ATLAS MC samples used in the wavelet performance study.

We need to scale the different samples used according to their cross section σ, so as to obtain reliable figures for the rates of QCD and boson jets at different $p_\perp$ scales. We choose a target integrated luminosity of $\mathcal{L} \equiv \int \mathcal{L} \, \mathrm{d}t = 5\,\mathrm{fb}^{-1}$, corresponding to a conservative estimate for the data collection during the first year of the LHC Run 2, and assign a weight of

$$w = \frac{\sigma \times \mathrm{BR} \times \mathcal{L} \times \varepsilon_{\mathrm{filter}}}{N_{\mathrm{gen.}}} \qquad (4.1)$$

to each event, in addition to any weight provided by the generator, where $N_{\mathrm{gen.}}$ is the number of generated events in the full sample.

Since we found in Sec. 3.4s, that wavelet-based cleaning yields improvements over standard methods mainly at moderately high pile-up multiplicities, and since mc14 samples "only" contain events with actual number of interactions $N_{\mathrm{PU}} \leq 40$, we have chosen to study only events with N_{PU} distributed uniformly on the interval [30, 40]. This allows us to study the impact of our wavelet analysis under conditions where this impact is expected to be most pronounced.

As mentioned, our samples are in the xAOD format, and so we must decide on what information to use, in order to perform a study which is comparable to that performed in toy simulation, cf. Secs.3.1–3.4. Since we are not particularly concerned with muons, we will be using objects from the xAOD::InDetTrackParticles collection as the basis for our track-based wavelet analysis. (Some of) these tracks are associated to objects in the xAOD::PrimaryVertices collection, which we will use to distinguish hard scatter (HS) tracks from pile-up (PU) tracks. As described in Sec. 3.1, the track-based methods necessitates a correction of the deflection in ϕ of charged

particles in the detector. Therefore we will base our jet clustering on the objects in the xAOD::chargedJetETMissPFO_eflowRec and xAOD::neutralJetETMissPFO_eflowRec containers, which hold the charged and neutral particle flow (PFlow) objects, resp., reconstructed using the eFlowRec algorithm described in Sec. 1.4. Finally, we will be needing the xAOD::TruthParticle container to identify the *W* jets by matching them in dR to the truth level *W* boson.

As in the toy simulation, jets are clustered using the FASTJET library [77], due to its ease of configuration, and since the official ATLAS JETREC tool didn't provide any additional corrections at the time of writing.

■ Kinematic distributions

The processes specified in Tab. 4.1 allows us to study the impact of wavelet analysis on boosted jet mass reconstruction, but not on jet energy resolution, at least not in the way proposed in Sec. 3.1. This is because the xAOD EDM in the available samples do not contain generator-level information about pile-up particles, and since we do not have access to the hard event in a pile-up free setting, we cannot compute the ratio in Eq. (3.1). However, based on the findings in Sec. 3.4, we are confident that the following results for the boson jet mass should also be representative for the jet energy resolution.

The chosen samples allow us to study the impact of wavelet methods over a large $p_\perp$ range. To better understand our samples, Fig. 4.1 shows the $p_\perp$ and mass distributions for jets matched to the truth-level *W* boson in the HVT(X TeV) / $WW_{(X\,\mathrm{TeV})} \to \ell\nu q\bar{q}$ samples.
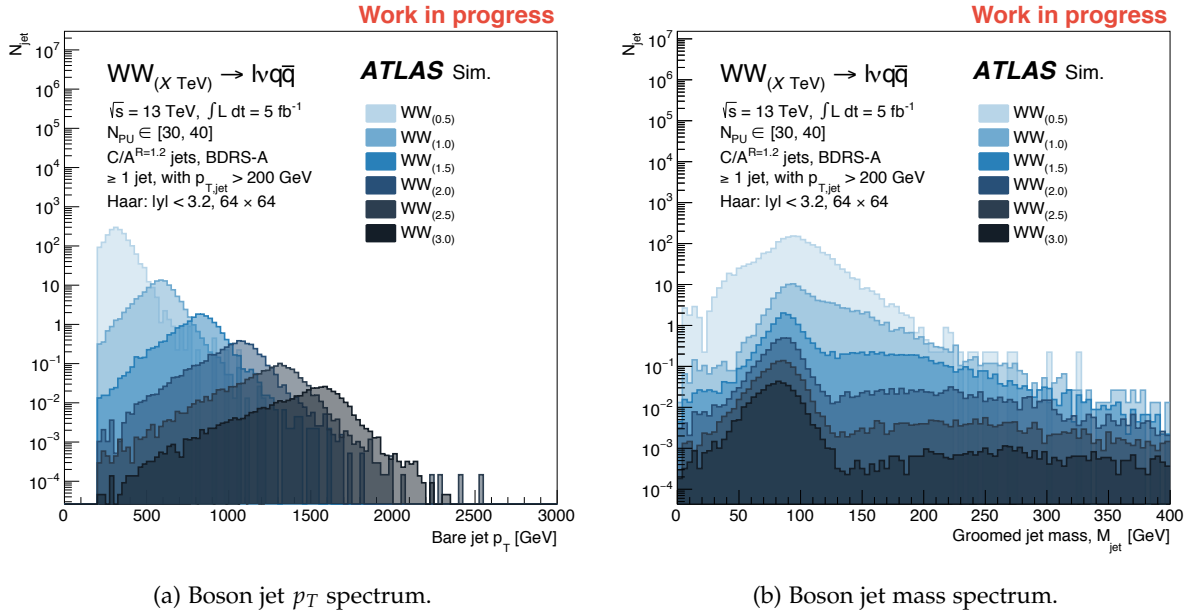
(a) Boson jet p_T spectrum.

(b) Boson jet mass spectrum.

Figure 4.1: Distributions of boson jet $p_\perp$ (left) and mass (right) in the ATLAS MC signal samples, cf. Tab. 4.1.

From Fig. 4.1(a) we see how the number of jets falls of rapidly, reflecting the cross sections for the different signal resonance mass points, cf. Tab. 4.1: at $\mathcal{L} = 5\,\mathrm{fb}^{-1}$ we expect $\approx$ 1.5 HVT(3.0 TeV) events in *total*. We also see how the peak in the $p_\perp$ distributions

grows linearly as $\sim M_{WW}/2$, modulo the roughly constant offset introduced by the pile-up collisions.

Fig. 4.1(b) shows the mass distribution for boson jets, exhibiting a clear peak at $M_{\text{jet}} \sim M_W \approx 80\,\text{GeV}$, as we had hoped. Also since boson jets from the high-mass resonance will have higher $p_\perp$, cf. Fig. 4.1(a), this results in naturally sharper mass peaks, for the reasons noted in Sec. 3.4. We also note, that the tail extending towards large masses, for all resonance samples, is suppressed by orders of magnitude (observe the logarithm scale).

Fig. 4.2 shows the same distributions for jets in the ATLAS QCD dijet samples.
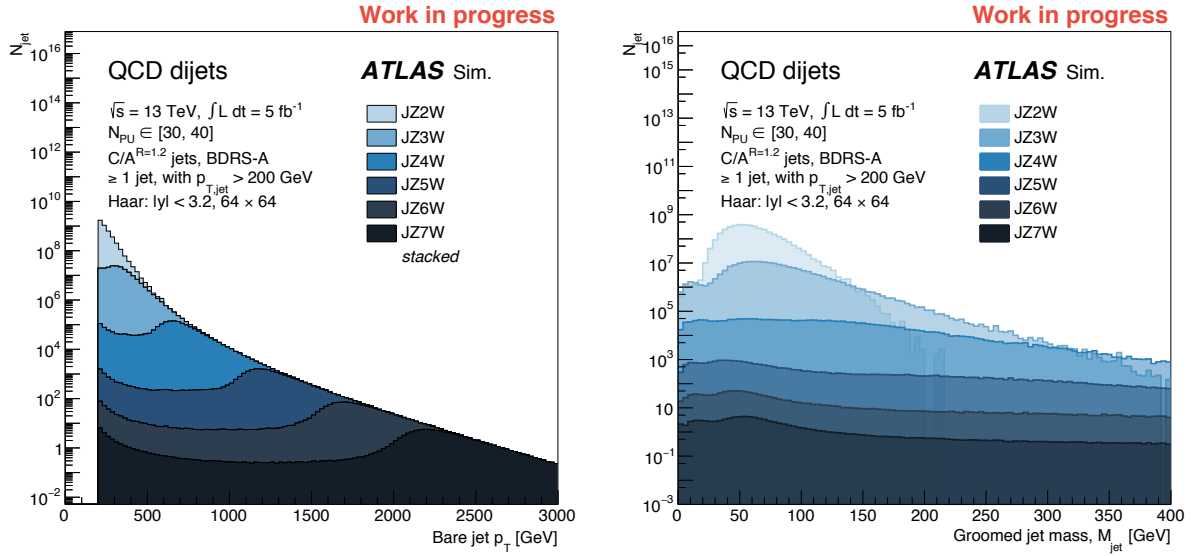
(a) QCD jet p_T spectrum.

(b) QCD jet mass spectrum.

Fig. 4.2(a) shows the stacked distribution of QCD jet $p_\perp$ for the six different JZxW samples used, cf. Tab. 4.1. We see the power-law like structure of the spectrum, as expected for non-resonant jets, deriving from the leading order $2 \to 2$ matrix element in Eq. (1.42) as well as the shape of the PDFs. Also, it is clear why creating separate $p_\perp$ slices is necessary for gaining sufficient statistics in the tail of the distribution.

Fig. 4.2(b) shows the mass spectrum, which shows no signs of consistent, resonant structure the way Fig. 4.1(b) does. However, the all slices have broad peaks at $M_{\text{jet}} \sim 50 - 100\,\text{GeV}$, with a tail towards larger M_{jet} which extends further for larger $p_\perp$, with an upper limit of $M_{\text{jet}} \lesssim \hat{p}_{\perp\text{max}}$ for each slice. The dip at $M_{\text{jet}} \approx 30 - 40\,\text{GeV}$ is a consequence of the $p_{\perp\text{jet}} > 200\text{GeV}$ cut imposed on all reconstructed, BDRS-A groomed jets.

Figure 4.2: Distributions of QCD jet $p_\perp$ (left) and mass (right) in the AT-LAS MC signal samples, cf. Tab. 4.1.

■ Track assignments

The track-based wavelet methods presented in Sec. 2.3 rely on our ability to (1) reconstruct charged particle tracks in the detector and (2) to correctly tell whether these originate from the hard scatter

process or from pile-up. In the toy model simulations, making this distinction was easy: we used generator level information to make this distinction, thus enabling a 100% track-vertex matching efficiency. However, faced with the imperfections of actual, physical detectors, we are forced to make some choices.

In official ATLAS MC, in the xAOD format, we have the following three categories of tracks:

(1) tracks, which we are positively sure belong to the primary vertex (those associated to the first object in xAOD::PrimaryVertices);

(2) tracks, which we are positively sure belong to a secondary vertex (those associated to any other object in xAOD::PrimaryVertices);

(3) tracks, which we cannot, initially, say come from either the primary or any of the secondary vertices (that is, the objects in xAOD::InDetTrackParticles which are not associated to any of the objects in xAOD::PrimaryVertices).

Fig. 4.3 shows the $p_\perp$ distributions of tracks in either category for a particular signal sample.

Figure 4.3: Track $p_\perp$ distributions depending on the track category, depending on whether the track is associated to the primary/hard scatter vertex (cat. 1; red), to another (pile-up) vertex (cat. 2; light blue), or to no vertex at all (cat. 3; dark blue).
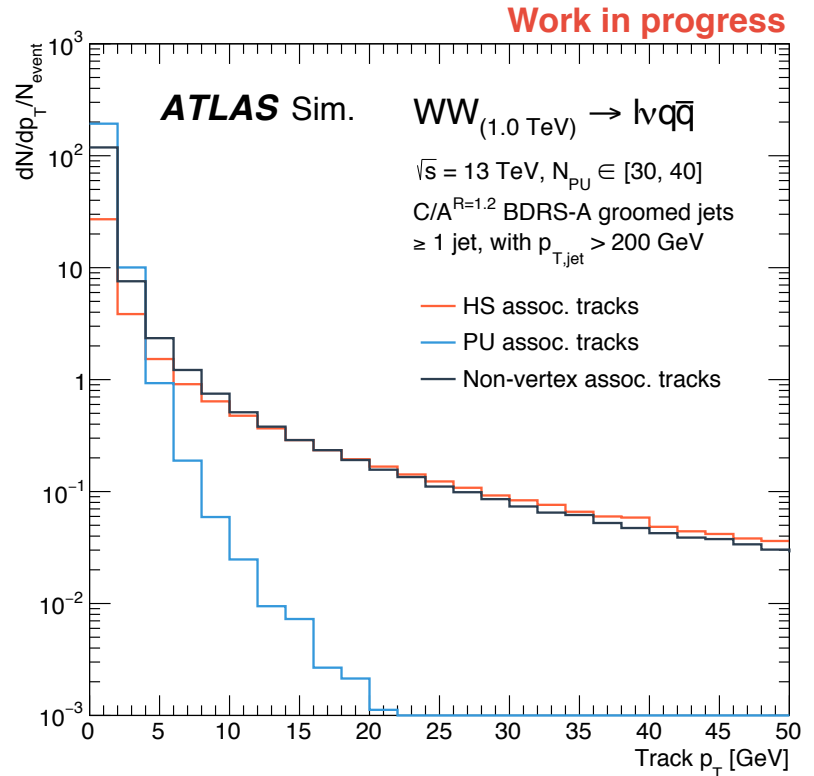
Here we see three things: first, there is a clear difference between the $p_\perp$ spectra for hard scatter (HS; cat. 1) and pile-up (PU; cat. 2) tracks; second, the number of tracks not associated to any primary vertex (cat. 3) is highly non-negligible; and third, this category clearly contain ostensibly hard scatter (from convergence to the hard scatter spectrum at large $p_\perp$) and pile-up (from increase at low $p_\perp$) components.

In order to use the track-based wavelet methods, we have to decide which tracks we want to assign to the hard scatter process and which ones we want to assign to pile-up. If cat. 3 above had been empty, the assignment would have been easy: assign all tracks in cat. 1 to the hard scatter process and all those in cat. 2 to pile-up. But since cat. 3 in non-empty (due to poor track quality and impact parameter errors), we need to decide on what to do with its contents. We risk making two types of errors:

Type 1 (false positive) error. Labeling a pile-up track as coming from the hard scatter process. The fraction of such errors is called α.

Type 2 (false negative) error. Labeling a hard scatter track as coming from a pile-up interaction. The fraction of such errors is called β.

Type 1 errors are not so dangerous: in the limit of labeling *all* tracks as coming from the hard scatter process, the track-based wavelet methods just have no effect, their performance increasing with decreasing α, smoothly recovering optimal performance for $\alpha = 0$. However, type 2 errors have the potential to be disastrous: if we incorrectly label a hard scatter track as coming from pile-up, it actively has the effect of killing the corresponding wavelet component, to which we would believe that the hard scatter contributes significantly, cf. Sec. 2.3. Therefore, while type 1 errors "only" deteriorates performance, type 2 errors actively kill parts of the event that we would want to keep.

Based on this reasoning, we decide to be as conservative as possible, and only label the tracks in category (2) as coming from pile-up and label the tracks in both categories (1) and (3) as coming from the hard scatter process. In this way, we are certain that we are not committing any type 2 errors, while we accept making type 1 errors and gladly pay the price of sub-optimal performance.

■ Comparison with toy MC

Having settled on a way to distinguish hard scatter tracks from pile-up ones, the first step in assessing the effect of wavelet methods in ATLAS MC is by studying the coefficient distributions. This is done in Fig. 4.4. These plots show the distributions in ATLAS MC compared to those in toy MC. The physics processes are $HVT(1.0\,\mathrm{GeV}) \to \ell\nu q\bar{q}$ and $W(\to q\bar{q}) + \mathrm{jets}$, the pile-up multiplicities ($N_{\mathrm{PU}} \in [30, 40]$, and $\langle \mu \rangle = 40$) as well as jet $p_\perp$ cuts (200 and 300/,GeV), for ATLAS and toy MC, resp., all differ slightly, but should allow for a general comparison. The RMS'es for the distributions in each bin is plotted, along with indicative FDN and SDN thresholds.

We see the same features noted in Sec. 3.2: the $(m_y, m_\phi) = (0, 0)$ coefficient is special, being always positive and having a mean of $\sim 25\,\mathrm{GeV}$; the $(0, 1)$ coefficient, albeit having a relatively much larger RMS than in toy MC, still reflects transverse momentum conservation, the difference in RMS being attributable to the simplicity of our toy model, cf. Sec. 3.1; the $(2, 0)$ coefficient exhibits the same sensitivity to the pseudorapidity plateau structure; the remaining scales exhibit

(a) ATLAS MC: HVT$(1.0\,\text{GeV}) \rightarrow \ell\nu q\bar{q}$.

(b) Toy MC: $W(\rightarrow q\bar{q})$ + jets.
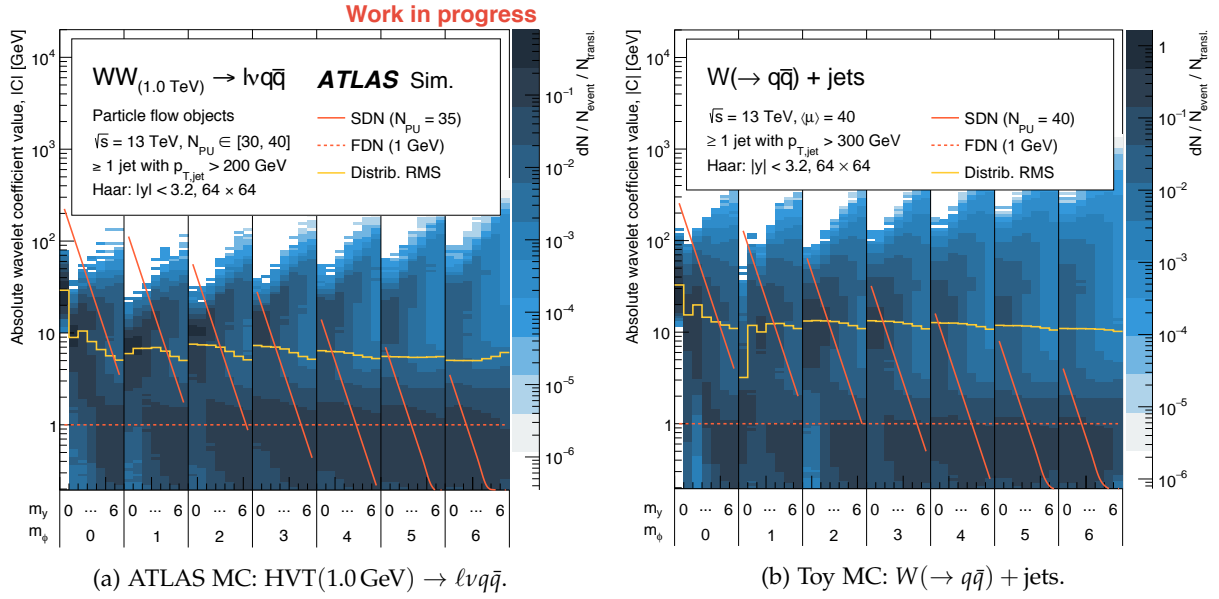
Figure 4.4: Coefficient value distributions in full events, showing the same as Figs. 3.8 and 3.9, for ATLAS MC (left) and toy simulation (right; see text for details). Events have slightly differing pile-up multiplicities ($N_{\text{PU}} \in [30, 40]$ and $\langle\mu\rangle = 40$, resp.) and jet $p_\perp$ cuts (200 and 300 GeV, resp.). The rise in the RMS'es in ATLAS MC for large $m_{y,\phi}$ is due to the zero-valued coefficients erroneously not being included in the RMS computation, and therefore it holds no physical significance.

(a) ATLAS MC: $WW_{(1.0\,\text{GeV})} \rightarrow \ell\nu q\bar{q}$.

(b) Toy MC: $W(\rightarrow q\bar{q})$ + jets.
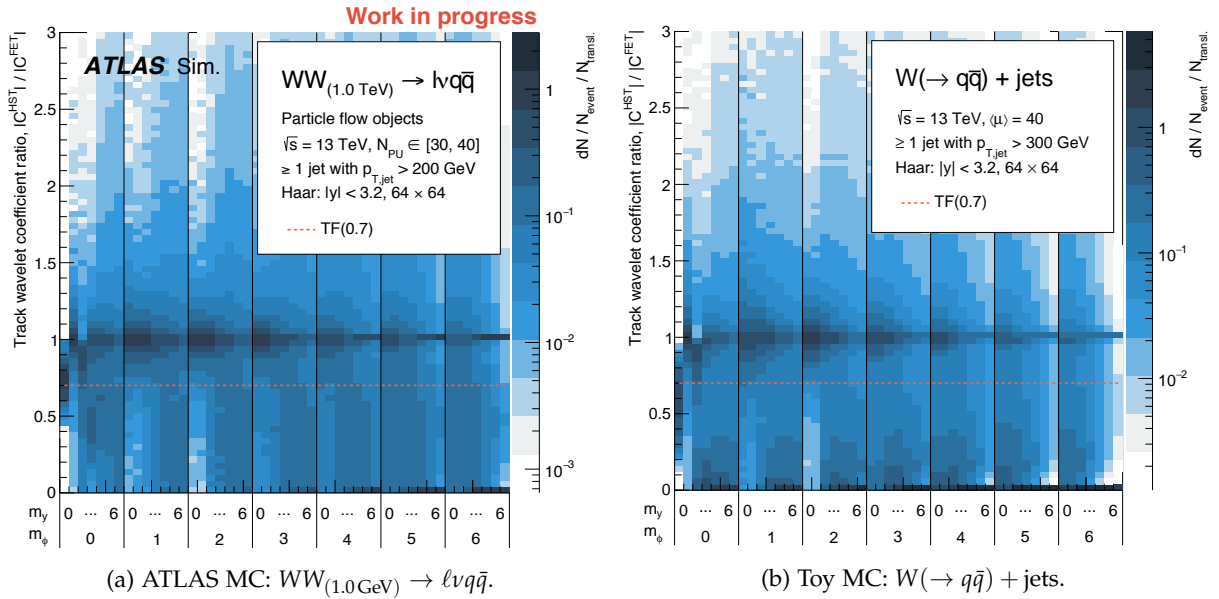
Figure 4.5: The distribution of ratios of same-indiced wavelet coefficients, for hard scatter tracks only to those for tracks in the full event, in ATLAS (left) and toy MC (right; see text for details). A typical track filtering threshold also shown (0.7; dashed red).

the same, regular behaviour as seen in Figs. 3.8 and 3.9. Similarly, the shapes of the distributions in each bin are similar, although this is not as easily quantifiable, since the distributions represent different processes and are subject to different conditions.

The coefficient distributions give no nasty surprises, so we turn the the distributions of the ratios of track coefficient, cf. Eq. 2.46. This is shown in Fig. 4.5 for both ATLAS and toy MC, under the same conditions as Fig. 4.4.

Again, we see similar behaviours for the two setups: the ratio for $(m_y, m_\phi) = (0,0)$ is confined to be ≤ 1, since the set of hard scatter tracks in a subset of the set of all tracks; the distribution for $(0,0)$ is shifted upwards in ATLAS MC, relative to toy MC, due to our conservative track assignment above, labeling ostensible pile-up track as coming from the HS, thereby increasing the ratio on the right hand side of Eq. 2.46; the ratio for $(0,2)$ is sensitive to the ATLAS ID $|\eta| < 2.5$ range, resulting in a distribution which peaks and sharply drops off just below 1; the remaining scales exhibit the same, regular behaviour as in toy MC. Notice that there is a marked excess of cases with ratios of zero, across all scales, for toy compared to ATLAS MC. This can again be attributed to our conservative track assignment.

Similarly to the coefficient distributions in Fig. 4.4, the track coefficient ratio distributions in Fig. 4.5 appear to have similar shapes in both sets of simulation. However, Fig. 4.5 displays distributions of *ratios* of coefficients, rather than of absolute values as in Fig. 4.4. Therefore, we consider it more appropriate perform a more detailed comparison of the distribution shapes in Fig. 4.5, since here many of the differences between the toy and ATLAS MC samples ought to divide out. Fig. 4.6 shows the $(m_y, m_\phi) = (3,3)$ slices of Fig. 4.5 for both ATLAS and toy MC.
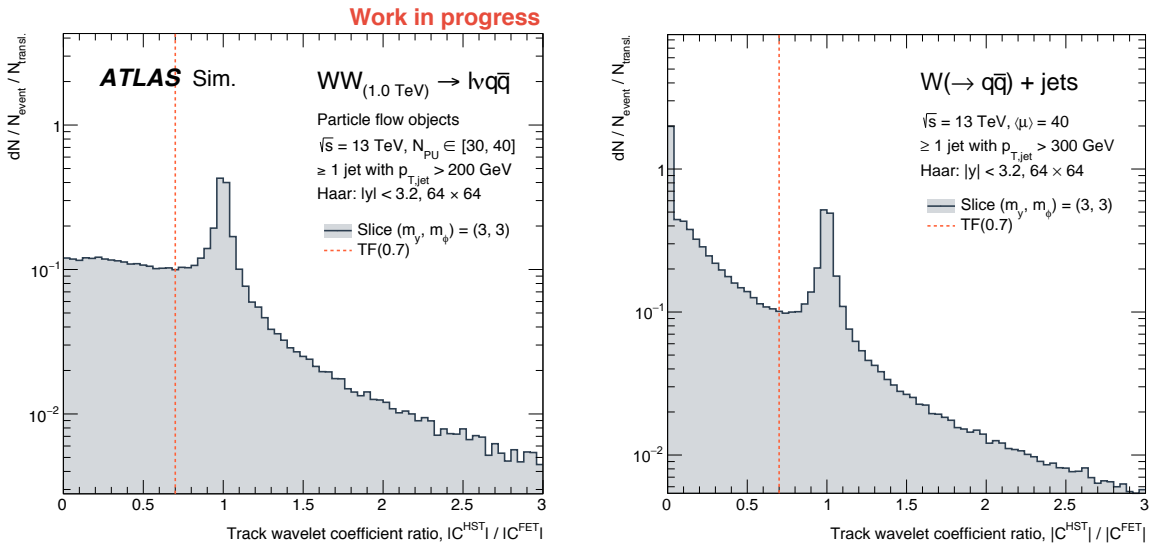
(a) ATLAS MC: $WW_{(1.0\,\text{GeV})} \to \ell\nu q\bar{q}$.

(b) Toy MC: $W(\to q\bar{q}) + \text{jets}$.

Here we see the same smooth, power-law spectrum as noted in Sec. 3.2, with prominent peaks at 1, and similar widths for the two setups. Finally, we note the excess of low and zero-valued ratios in toy MC compared to ATLAS, as described above.

A final check may be looking at the output-to-input ratio of particle momenta after applying the TF(0.7) method. This is done in Fig. 4.7.

Figure 4.6: Slices at scales $(m_y, m_\phi) = (3,3)$ of Fig. 4.5 for ATLAS (left) and toy MC (see text for details). A typical TF threshold is also shown (0.7; dashed red).

(a) ATLAS MC: $WW_{(1.0\,\mathrm{GeV})} \to \ell\nu q\bar{q}$.

(b) Toy MC: $W(\to q\bar{q}) + \mathrm{jets}$.
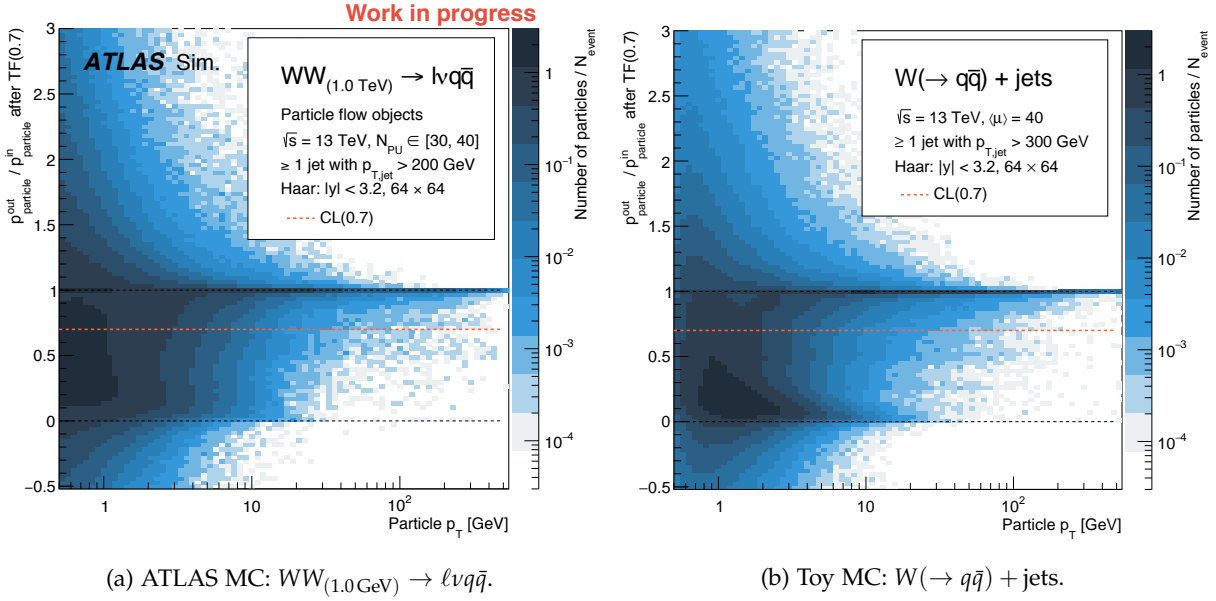
Figure 4.7: Output-to-input particle $p_\perp$ ratios, after application of the TF(0.7) method cf. Fig. 2.6, as a function of the input particle $p_\perp$, for ATLAS (left) and toy MC (right; see text for details). The CL(0.7) cut (dashed red) is shown.

We see that the distributions are largely similar: tails at low $p_\perp$ extending out to ratios above one and below zero; both distributions have a prominent valley-structure, consistent with the separation of hard scatter particles from pile-up, with ridges forming at ratio up to one and down to zero; moderate to high-$p_\perp$ ($\gtrsim 10\,\mathrm{GeV}$) particles are sharply peaked at ratios of one, and are thus only modified to a vanishing extent; we see an excess in particles with low $p_\perp$ ($\sim 1\,\mathrm{GeV}$) and low ratios ($\sim 0 - 0.5$) in toy MC, which can be attributed to the larger amount of pile-up as well as the optimal track-vertex–matching in toy MC.

Generally we find, that the distributions of coefficients, their ratios (in tracks), and the particles discarded by the particle cleaning method, behave similarly in ATLAS MC as in the toy simulation covered in Secs. 3.1 to 3.4. This indicates that the methods and their results should generally carry over as well and, in particular, that the chosen wavelet analysis structure and the optimal parameter configuration, TF(0.7) $\oplus$ CL(0.7), should also perform well when presented with more realistic events. Therefore we choose to keep using this wavelet analysis, to facility comparison with the results of Sec. 3.4 in order to assess whether–and if so, why–the performance differs. However, in the interest of thoroughness, one might perform a separate optimisation for official ATLAS samples.

■ Boosted boson jet mass

We are now confident that the basis of the our wavelet analysis is sound, also in ATLAS MC, and we can turn to studying the impact of our wavelet analysis on the reconstruction of the mass of boosted boson jets. Seeing as we are limited to studying low-to-moderate $\langle\mu\rangle$, we focus on configurations with $N_{\mathrm{PU}} \in [30, 40]$. The $p_\perp$ dependence is accessible to us through the different signal resonance mass points,

cf. Fig. 4.1(a). Jets are clustered in PFlow objects with $p_{\perp\text{jet}} > 200\,\text{GeV}$ and for each mass point we choose a $p_{\perp\text{jet}}$ interval of $\pm 50\,\text{GeV}$, for $WW_{(0.5\,\text{TeV})}$, or $\pm 100\,\text{GeV}$, for the remaining mass points, around the jet $p_{\perp}$ peak for the corresponding signal sample, and compare boson jets to QCD jets within this interval, in order to study jets which are kinematically similar. We then apply the $\text{TF}(0.7) \oplus \text{CL}(0.7)$ wavelet analysis to the PFlow objects, cluster jets in the scaled and slimmed output collection (with a lower $p_{\perp}$ threshold), and keep those matched within $dR < 0.6$ to jets in the original event. The distribution of jet mass reconstructed in this way is shown in Fig. 4.8 for the limiting cases of $WW_{(0.5\,\text{TeV})}$ and $WW_{(3.0\,\text{TeV})}$, resp. In both figures, we see that the W-jet distributions have been scaled up factors of 2.0×10^{6} and 4.4×10^{3} to compensate for the vast difference in cross sections, cf. Tab. 4.1.

(a) $M_{WW} = 0.5\,\text{TeV}$.

(b) $M_{WW} = 3.0\,\text{TeV}$.
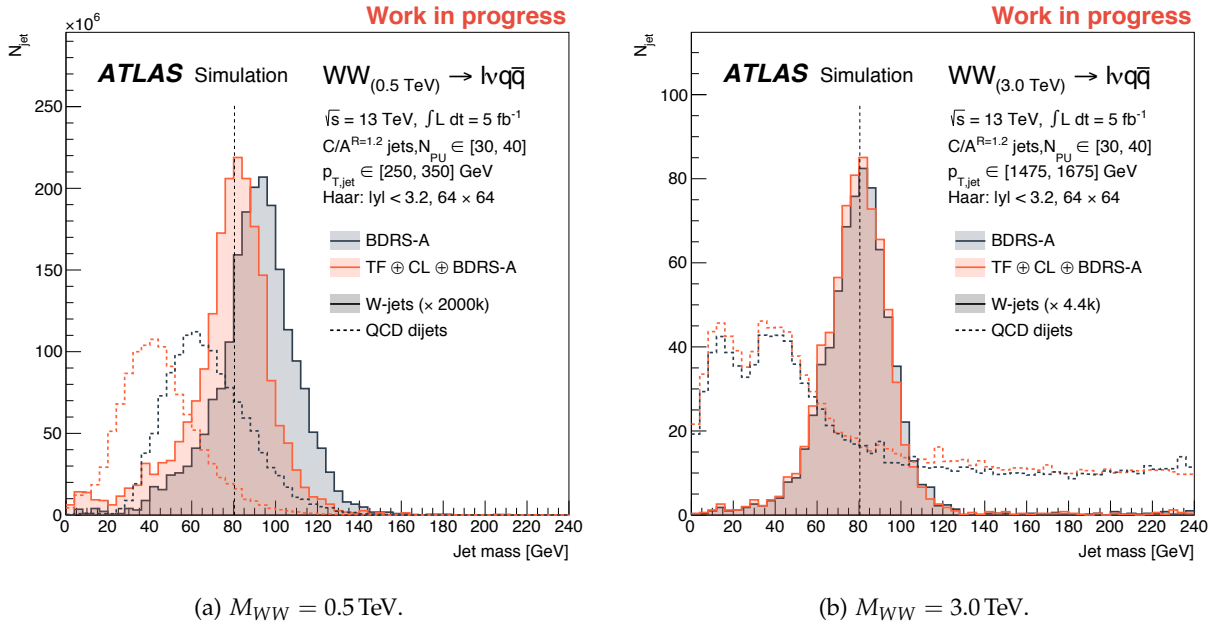
From Fig. 4.8(a) we see that the wavelet analysis has the effect of narrowing the W-jet mass peak slightly and removing its bias, as well as shifting the QCD jet distribution towards lower M_{jet} values, effectively removing jets from the area around the W boson pole mass. This is the same behaviour found in e.g. Fig. 3.28(a). However, we also notice a few differences. First, the shape of the BDRS-A groomed distribution is not as broad and as skewed towards larger M_{jet} values in ATLAS MC as in toy. This can be explained both by the lower number of pile-up collision and by the fact that the toy model smearing described in Sec. 3.1 is too simplistic, and the reason we used two times the HCAL resolution in the smearing was exactly to obtain a "worse" jet mass resolution. Second, we see that there is no tail at $M_{\text{jet}} \approx 0\,\text{GeV}$ in ATLAS MC. Despite the reason for this difference not being immediately clear, the effect occurs at low mass, and thus has no effect on our findings. Moreover, this effect is present also for BDRS-A grooming alone and is therefore an effect of the

Figure 4.8: Jet mass distributions for boosted W boson jets (filled) compared with QCD jets (dashed). Distributions for BDRS-A groomed jets (dark blue) and groomed jets clustered in the $\text{TF}(0.7) \oplus \text{CL}(0.7)$ wavelet-cleaned event (red) are shown. The dashed black line marks the W boson pole mass.

grooming procedure, or the input objects, rather than of the wavelet analysis. Finally, we note that the improvement found here, while evident, are less pronounced than in the toy MC, cf. Fig. 3.28(a). This, however, is exactly what we expected from the conservative track assignment introduced previously.

From Fig. 4.8(b) we draw the same conclusion as from Fig. 3.29: at high $p_\perp$, the wavelet methods provide no further sharpening of the jet mass peak, at least not at moderately low $\langle \mu \rangle$. However, in contrast to the findings for the toy MC, we see that at high $p_\perp$, while the distribution bias is gone, its not narrower than at $p_{\perp\text{jet}} \approx 300 \,\text{TeV}$. This is no cause for despair, since our previous smearing was deceptively simple, being perfectly gaussian and having just simple $p_\perp$ dependencies through Tab. 1.4. The ATLAS simulation infrastructure passes the events (after decay and hadronisation) through a GEANT4 model of the ATLAS detector, as noted above, which results in an immensely more realistic—and involved—estimate of the actual measurements in the ATLAS detector. Therefore, we interpret the distribution width in Fig. 4.8(b) as a result of the this more complicated smearing, and only note that the conclusions regarding the impact of the wavelet analysis hold for both toy and ATLAS MC.

Finally, we want to study the dependence of our boson jet mass metric on the jet $p_\perp$. As in Sec. 3.4, we use the $S/\sqrt{B}$ within a fixed $M_{\text{jet}} \in [60, 100]\,\text{GeV}$ as our figure of merit, since it is sensitive to both the width and the bias of the jet mass distribution, as well as to the effectiveness of the wavelet analysis in discarding QCD jets around the W pole mass. We compute the $S/\sqrt{B}$ value for jets within each of the $p_{\perp\text{jet}}$ intervals mentioned above, for groomed jets clustered in the original event and for groomed jets clustered in the wavelet-cleaned event. The resulting ratios are shown in Fig. 4.9.

Figure 4.9: Ratio of $S/\sqrt{B}$ for M_{jet} distributions after wavelet analysis and jet grooming to those after only jet grooming, for boson vs. QCD jets, computed in the fixed $M_{\text{jet}} \in [60, 100]\,\text{GeV}$ interval, as function of the jet $p_\perp$ for $N_{\text{PU}} \in [30, 40]$.
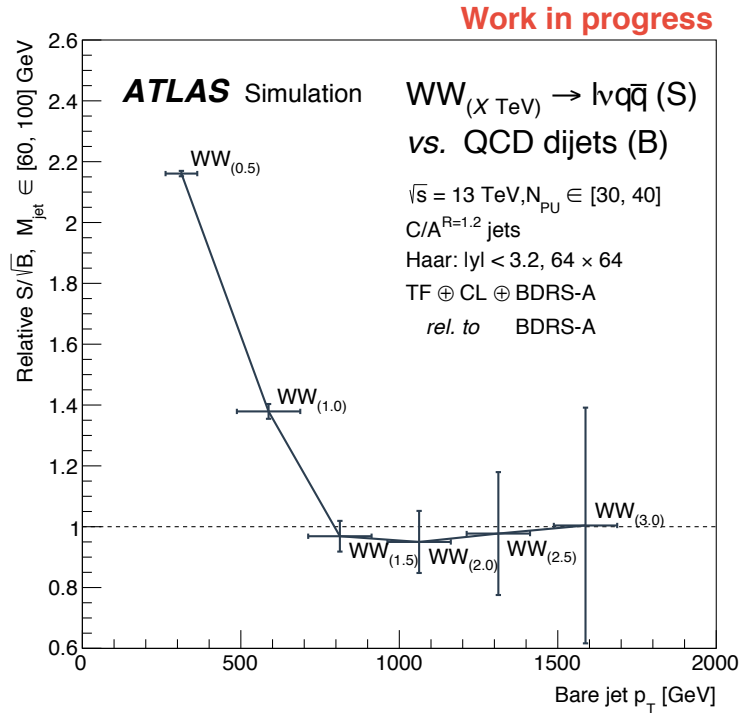
Here, each point on the graph is computed for the indicated signal resonance mass point. The horizontal error bar shows the $p_{\perp\text{jet}}$ interval used in the computation. The vertical error bar shows the statistical uncertainty on the $S/\sqrt{B}$ ratio, computed by Eq. (3.4)

The plot shows the same general behaviour as that found in Fig. 3.30: for jets with moderately low $p_\perp$ we see a clear improvement in our ability to distinguish boson jets from their non-resonant counterparts; but for increasing $p_\perp$ this improvement disappears quickly, returning to the baseline performance for standard jet grooming. The main point of the error bars is to show, that any dip below the unit ratio is within expected fluctuations[2]. In regards to the absolute values of these improvements, and the position of the "kink" where the improvement vanishes, the graph in Fig. 4.9 is seen to correspond closely to the $\langle\mu\rangle = 40$ contour in Fig. 3.30. This is an indication that, despite the differences their closeness to realism, final results for toy and ATLAS MC are comparable. This also means that, although we cannot presently study the behaviour of the jet mass improvement with $\langle\mu\rangle$ in ATLAS MC, nor study any improvements in jet energy resolution in the way proposed in Sec. 3.1, we can now have some confidence that, if we had been able to do so, the findings would have been similar.

This concludes our discussion of the general application and performance of wavelet-based methods in hadron collision events. We have proposed a set of wavelet methods, discussed their merits, optimised one full wavelet analysis, and studied its impact on our ability to reconstruct jet energies and masses reliably. Now, we want to lose our training wheels and see if—and to what degree—wavelet analyses can aid searched for new physics.

[2] The assumption of no correlation going into Eq. (3.4) is arguably quite a stretch. Additionally, the S and B figures are computed based on the counts, scaled by the weight in Eq. (4.1), rather than the *actual number* of jets. Therefore, the error bars in Fig. 4.9 are only intended to be indicative or the rough scale of the expected fluctuations as they would appear in data.

4.2 Diboson resonance search

Back in Secs. 1.1 and 2.2 we laid out the motivation for this study: we want to improve jet mass measurements in order to gaing sensitivity in searches for new physics diboson final states. In the preceding sections we have thoroughly studied wavelet methods for hardon collision in general terms and—we hope—showed that they hold great promise. In this section, we therefore present a "mock" physics analysis, specifically a search for narrow resonances in the $\ell\nu q\bar{q}$ final state. We present the samples, the definition of physics objects, the event selection, and, finally, the impact of our optimised wavelet analysis in the search sensitivity.

■ Samples

We noted in Sec. 1.1 that we would be using the $WW \to \ell\nu q\bar{q}$ as our benchmark search. Therefore, this section largely follows the ATLAS Exotics Diboson $\ell\nu q\bar{q}$ Run 1 analysis [16], with only minor changes to the samples used, the object definition, and event selection. This is done to ensure that our results will be as close to a "real" analysis as possible considering, while considering the limited scope of this study.

Seeing as out wavelet analysis methods have developed with an eye on the future, we choose to use samples with $\sqrt{s} = 13\,\mathrm{TeV}$, appropriate for the LHC Run 2. Therefore, due to the timing of this study, we are not been able to include actual data in this study in any meaningful way, so the following will be a purely simulation-based feasibility study.

Benchmark signal samples are used to study the expected sensitivity of the search, and we use the HVT samples introduced in Sec. 4.1. Since this is just a feasibility study, we only consider three of the major backgrounds found in [16]: $W \to \ell\nu + X$ ('W + jets'), $t\bar{t} \to \ell + X$ ('$t\bar{t}$'), and $Z \to \ell\ell + X$ ('Z + jets'). Observe the difference between the W/Z + jets samples used here (leptonically decaying vector boson) and those used in Secs. 3.1 through 3.4 (hadronically decaying vector boson). These are all three reducible backgrounds[3] for the following reasons:

W + jets. Produced through the diagrams in Fig. 3.3. Contains a true, leptonically decaying W boson, which might be selected by the analysis, provided it has sufficient $p_\perp$. The recoil jet will be a non-resonant jet ("QCD jet") which may be tagged as a hadronically decaying W boson, provided its mass falls sufficiently close to the W boson pole mass. Using the methods developed throughout this chapter, we hope to be able to reduce the contribution from mistaking such QCD jets for actual boson jets by improving the jet mass resolution.

t̄t. Produced through diagrams such as that in Fig. 4.10. In the semi-leptonic decay mode (with which we're concerned), it contains both a true leptonically decaying W boson as well as a true hadronically decaying W. What distinguishes it from the signal final state is

[3] The irreducible background, continuum WW production, was not available in the mc14 project despite it being the second largest background contribution in the boosted regime [16], see below. Nevertheless, as we are interested in killing non-resonant QCD jets, which are present in the major *reducible* background, W + jets, but not in continuum WW, we deem this compromise to be acceptable.

large jet multiplicity, in particular the presence of two *b*-jets.

Z + jets. Produced through the diagrams in Fig. 3.3. Contains a leptonically decaying *Z* boson, where one of the decay products may be tagged as the *W* boson lepton, provided the other lepton is not properly identified. Calorimeter (jet) energy mis-measurements may introduce an artificial missing energy, which may be reconstructed as the *W* boson neutrino. The recoil jet will, similarly to *W* + jets, be a QCD jet which may be tagged as a hadronically decaying *W* boson.

Figure 4.10: An example of $t\bar{t}$ production by gluon fusion, and semi-leptonic decay.

The samples used to model these processes have all been generated under the mc14 simulation project, processed in the xAOD format with the *_r5787_r5853 reco. tag. All samples, signal and background, are passed through a GEANT4 model of the ATLAS detector in order to realistically model the response of the various detector components.

The *W* and *Z* + jets samples are all generated in particle level $\hat{p}_\perp$ slices, from $[70, 140]\,\mathrm{GeV}$ to $> 500\,\mathrm{GeV}$, with SHERPA using the CT10 PDF set [85].

The $t\bar{t}$ samples are generated in POWHEG [86], with no kinematic segmenting but requiring at least one leptonically decaying *W* boson. Subsequent parton showering and hadronisation is performed in PYTHIA using the Perugia2012 tune [87]. The CTEQ6L1 LO PDF set [82] is used.

Detailed sample lists, specifying the datasets used, their ID, decay mode, kinematic cuts, cross section times branching ratio ($\sigma \times$ BR), generator filter efficiency, and number of generated events can be found in App. A.3, for both signal (Tab. A.1) and background (Tabs. A.2 and A.3) samples.

Throughout this mock analysis no pile-up re-weighting is performed, and consequently we are using the standard pile-up multiplicity profile present in mc14 xAOD datasets: a linear increase from $\approx 0.5\%$ to $\approx 3.2\%$ for $0 \leq N_{\mathrm{PU}} \leq 20$ followed by a plateau for $20 \leq N_{\mathrm{PU}} \leq 40$. This results in an average pile-up multiplicity of $\langle \mu \rangle \approx 25$ which is lower than nominal LHC Run 2 predictions. This means that, as far as pile-up mitigation is concerned, the following represents a 'worst case'-scenario for the Run 2.

■ Object definition

In order to consistently specify our final state of interest, we must define the physics objects involved. Disregarding τ leptons, being notoriously hard to identify, we let $\ell = e, \mu$, which are both much easier to identify cf. Sec. 1.4. Considering the diboson resonance mass limits set by previous analyses [14, 15, 16] cf. Sec. 1.1, we will only be interested in cases where the hadronically decaying *W* boson has enough transverse momentum that the decay products will be sufficiently collimated, cf. Eq. (1.53), so as to be reconstructed as a single, large-*R* jet. Therefore, the $\ell \nu q \bar{q}$ final state will be constructed from a reconstructed electron or muon, missing transverse energy,
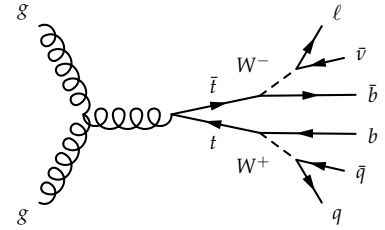
carried out of the detector by the neutrino, and (at least) one large-R jet. The object definition used in this study, closely following that in [16], is detailed below.

Leptons are selected as either 'signal' or 'veto' leptons, where the 'signal' object definition is tighter than the 'veto' definition and contained herein. This allows us to select events with one signal lepton, assumed to be from the leptonic W, and no (additional) veto leptons, thereby vetoing multi-lepton final states, like $Z(\to \ell\ell) + \text{jets}$.

Muons. All muons are required to be of type 'Combined', i.e. have tracks in both the ID and MS, cf. Sec. 1.4. Muons are required to pass the Muon Combined Performance (MCP) recommendations using the 'Loose' muon quality working point, as implemented in the `CP::MuonSelectionTool`. A cut on the longitudinal impact parameter $|\Delta z_0 \sin\theta| < 0.5\,\text{mm}$ (2.0 mm) is imposed for signal (veto) muons in order to keep only ones from the primary vertex (PV). Similarly, a cut on the transverse impact parameter significance $|d_0/\sigma_{d_0}| < 3.5$ is imposed for both signal and veto muons, to reject any muons produced at displaced vertices. Muons are required to be within $|\eta| < 2.5$ and have $p_\perp > 25\,\text{GeV}$ (20 GeV) for signal (veto) muons. The ratio of the scalar sum of $p_\perp$ of all tracks within a cone of $dR < 2.0$ around the muons, excluding the $p_\perp$ of the muon itself, to the muon $p_\perp$ must be $p_\perp^{\text{cone,20}}/p_\perp < 0.15$ for all muons. Additionally, the energy deposited in the calorimeter within a cone of $dR = 0.2$ around the muon, relative to the muon $p_\perp$, must be $E_\perp^{\text{cone,20}}/p_\perp < 0.14$ for all signal muons. Finally, the charge over momentum (q/p) measured in the ID and MS must agree within 5σ.

Electrons. No author requirement is (formally) imposed, since these are by construction always passed by the reconstructed electrons.[4] Signal (veto) electrons are identified using the 'Tight' ('Medium') menu available in the `xAOD::EgammaParameters::SelectionMenu`. As for muons, a $|\Delta z_0 \sin\theta| < 0.5\,\text{mm}$ (2.0 mm) cut is imposed for signal (veto) electrons, as well as a $|d_0/\sigma_{d_0}| < 6.0$ cut for both signal and veto electrons. Electrons are required to have $|\eta| < 1.37$ or $1.52 < |\eta| < 2.47$, i.e. be located in the LAr barrel or end-cap, excluding the crack region, cf. 1.3. Signal (veto) electrons are required to have $p_\perp > 25\,\text{GeV}$ (20 GeV). As for muons, no isolation cuts are imposed. Signal electrons are required to pass an object quality cut ('OTX cleaning': `xAOD::Egamma::isGoodOQ(1446) == true`). The ratio of the scalar sum of $p_\perp$ of all tracks within a cone of $dR < 2.0$ around the electron, excluding the $p_\perp$ of the electron itself, to the electron $E_\perp$ must be $p_\perp^{\text{cone,20}}/E_\perp < 0.15$ for all electrons. Additionally, the energy deposited in the calorimeter within a cone of $dR = 0.2$ around the electron, relative to the electron $E_\perp$, must be $E_\perp^{\text{cone,20}}/E_\perp < 0.14$ for all signal electrons. Finally, electrons are rejected if they overlap within $dR < 0.1$ of a veto muon, in order to avoid ambiguous definitions.

Missing $E_\perp$. The missing transverse energy used presently is taken from the MET_RefFinal[5] container, which uses the vector sum of energies associated to electrons, photons, taus, jets (anti-$k_\perp^{R=0.4}$ LCTopo, $p_\perp > 20\,\text{GeV}$), muons, and clusters not associated to physics objects. The contributions from each of the objects listed above are calibrated according to the prescriptions for the object type in question.

Jets. The initial event selection uses the standard CamKt12LCTopoJets jet collection, clustered with TopoClusters as input using the Cambridge-Aachen (C/A) algorithm with distance parameter $R = 1.2$, cf. Sec. 1.4. Note, that in the event selection, *no grooming* is applied to the jets. Jets are required to have $p_\perp > 30\,\text{GeV}$ and to be within $|\eta| < 2.0$. Finally, jets overlapping with a reconstructed veto electron within $dR < 0.8$ are discarded.

***b*-tagging.** In order to reject top events, see Fig. 4.10, we use a *b*-jet veto. This is done using the MV1 tagger, which is a neural network combination of the JetFitter, IP3D, and SV *b*-tagging algorithms in ATLAS, producing a single discriminating variable indicating the likeliness of a jet being initiated by a *b* quark. Anti-$k_\perp$ jets with a distance parameter of $R = 0.4$ are clustered using TopoClusters as input. Any one of these having a MV1 weight greater than 0.7892 is tagged as a '*b*-jet'.

The leading reconstructed fat jet will be identified as the hadronically decaying *W* bosons, and the four-vector sum of the reconstructed signal lepton and neutrino[6] is identified as the leptonically decaying *W* boson. The four-vector sum of these then allows us to compute the invariant mass of the diboson state, and this is where we will be looking for exotic diboson resonances.

■ Event selection

We use an event selection similar to that in [16], with a few exceptions: cuts that relate only to data events are not used (GRL, bad/corrupted events, hot tile veto); due to questions of whether trigger information is available in certain xAOD's and how to translate Run 1 trigger menus into Run 2 ones, it was chosen not to use trigger information; therefore, no trigger matching is performed either; jet mass and substructure cuts are not applied in the event selection, since we want to study the impact of wavelet-based methods on the jet mass spectrum, and since any substructure cuts should be (optimised and) applied only after the jet mass cuts. Additionally, as noted above, for the resonance masses of our primary interest, the decaying vector bosons will be sufficiently collimated that they may be reconstructed using a single, large-R jet collection (C/A$^{R=1.2}$). Therefore, we will be considering only the 'boosted regime' in [16], and not the 'resolved regimes', where the hadronically decaying bosons are reconstructed as two distinct small-R jets.

In the general event selection, we require: a reconstructed vertex

[5] See   http://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/MissingETUtility.

[6] See Sec. 1.4 and App. A.1.

with at least three associated tracks ('Primary vertex'); exactly one signal lepton and no additional veto leptons ('$N_{\text{lepton}} = 1$'); no anti-$k_\perp^{R=0.4}$ jets with $p_\perp > 20\,\text{GeV}$ and $|\eta| < 4.5$ which are flagged as 'BadLoose' by the JetSelectorTools package[7] ('$E_\perp^{\text{miss}}$ cleaning').

Additionally, specific to the boosted regime, we require: at least one reconstructed C/A$^{R=1.2}$ jet; that the leading C/A$^{R=1.2}$ jet ('signal jet') must have $p_\perp > 400\,\text{GeV}$ (we stress, that this cut is applied to the $p_\perp$ of the *un-groomed* jet); that the reconstructed, leptonically decaying vector boson (constructed from the signal lepton and the missing transverse energy) must have $p_\perp > 400\,\text{GeV}$; no b-jets outside $dR > 0.8$ of the signal jet ('b-jet veto') in order to discard top-events; a separation in azimuthal angle between the signal jet and the missing transverse energy of at least 1 ('$\Delta\phi(\text{J}, E_\perp^{\text{miss}}) > 1$') in order to avoid energy mis-measurements and therefore discard multi-jet events (almost exclusively in the e-channel); $E_\perp^{\text{miss}} > 30\,\text{GeV}$ in order to select mainly events with genuine missing energy from neutrinos, and discard events with $E_\perp^{\text{miss}}$ from energy mis-measurements.

Cutflow

A cutflow, including also the cuts not imposed in this study, is shown in Tab. 4.2 for the three considered background components.

Table 4.2: $\ell\nu q\bar{q}$ cutflow for the three considered background components.

Cut	Process			
	W + jets	$t\bar{t}$	Z + jets	Total
Preselection				
Total	2312307.	2054363.	279192.	4645862.
GRL[a]	2312307.	2054363.	279192.	4645862.
Detector problems[a]	2312307.	2054363.	279192.	4645862.
Primary vertex	2310296.	2054116.	278948.	4643360.
Trigger[b]	2310296.	2054116.	278948.	4643360.
$N_{\text{lepton}} = 1$	718308.	717648.	86534.	1522490.
Trigger matching[b]	718308.	717648.	86534.	1522490.
Jet cleaning[a]	718308.	717648.	86534.	1522490.
$E_\perp^{\text{miss}}$ cleaning	703479.	703736.	84623.	1491838.
Boosted regime				
≥ 1 C/A$^{R=1.2}$ jet	560597.	643829.	68489.	1272915.
$p_\perp(\text{J}) \geq 400\,\text{GeV}$	4327.	11774.	429.0	16529.
$p_\perp(\ell\nu) \geq 400\,\text{GeV}$	1484.	1785.	66.54	3335.
Jet mass[c]	1484.	1785.	66.54	3335.
Jet substructure[b]	1484.	1785.	66.54	3335.
b-jet veto	1337.	454.9	58.16	1850.
$\Delta\phi(\text{J}, E_\perp^{\text{miss}}) > 1$	1296.	436.0	51.81	1784.
$E_\perp^{\text{miss}} > 30\,\text{GeV}$	1240.	431.3	42.38	1713.

[a] Not relevant for the present study.
[b] Not considered for the present study.
[c] Omitted for later study.

We note the that total, raw number of $W(\to \ell\nu)/Z(\to \ell\ell) + X$ events refers to the number of events *used* in this study, rather than

the total expected number, since only samples with $\hat{p}_\perp > 70\,\text{GeV}$ are used. Nevertheless, the number of $W + \text{jets}$ events is 8.3 times larger than the number of $Z + \text{jets}$ events, despite the seeming similarity of the processes, see e.g. Fig. 3.3. However, this is consistent with the results from [89] and is expected for two reasons as noted in Sec. 1.1 (disregarding pole mass differences, the effect of which vanishes at large momentum transfer Q^2): first, inclusive W productions wins a factor of two over Z production, due to W being charged; and second, the W hadronic decay is CKM suppressed, and, unlike the Z boson, it has no invisible decay modes, thereby increasing the branching ratio of $\ell\nu$ decay modes (which is filtered in the present samples). This leads to an of roughly $\text{BR}(W \to \ell\nu)/\text{BR}(Z \to \ell) = 32.4\%/10.1\% \approx 3.2$, cf. Eqs. (1.17) and (1.18). This means that, from Tab. 4.2, $\sigma(pp \to W + X)/\sigma(pp \to Z + X) \approx 2.6$, which roughly agrees with the charge difference.

The final count ratios are approximately $29 : 10 : 1.0$, compared to $29 : 2.1 : 1.0$ for the ATLAS Run 1 $\ell\nu$J diboson analysis [16]. This difference (*after* $\ell\nu$J selection, modulo the omission of multiple cuts in the present event selection) is mainly due to the differences in cross-section at 8 and 13 TeV [30], where $\sigma(pp \to t\bar{t} + X)$ rises faster than $\sigma(pp \to W/Z + X)$ which, both rise at the same rate.

Tab. 4.3 shows the same cutflow for the six benchmark signal mass points used in this study.

Cut	Process (HVT)					
	0.5 TeV	1.0 TeV	1.5 TeV	2.0 TeV	2.5 TeV	3.0 TeV
Preselection						
Total	6673.	393.9	65.36	15.49	4.399	1.371
GRL[a]	6673.	393.9	65.36	15.49	4.399	1.371
Detector problems[a]	6673.	393.9	65.36	15.49	4.399	1.371
Primary vertex	6673.	393.9	65.36	15.49	4.399	1.371
Trigger[b]	6673.	393.9	65.36	15.49	4.399	1.371
$N_{\text{lepton}} = 1$	3994.	262.5	45.15	10.82	3.061	0.966
Trigger matching[b]	3994.	262.5	45.15	10.82	3.061	0.966
Jet cleaning[a]	3994.	262.5	45.15	10.82	3.061	0.966
$E_\perp^{\text{miss}}$ cleaning[c]	3919.	257.8	44.29	10.60	3.000	0.946
Boosted regime						
$\geq 1\ \text{C/A}^{R=1.2}$ jet	3517.	250.6	43.80	10.54	2.987	0.943
$p_\perp(\text{J}) \geq 400\,\text{GeV}$	32.70	178.5	41.66	10.35	2.960	0.938
$p_\perp(\ell\nu) \geq 400\,\text{GeV}$	8.675	153.4	40.94	10.30	2.952	0.936
Jet mass[c]	8.675	153.4	40.94	10.30	2.952	0.936
Jet substructure[b]	8.675	153.4	40.94	10.30	2.952	0.936
b-jet veto	8.008	144.5	38.42	9.625	2.750	0.872
$\Delta\phi(\text{J}, E_\perp^{\text{miss}}) > 1$	8.008	143.2	38.01	9.541	2.730	0.865
$E_\perp^{\text{miss}} > 30\,\text{GeV}$	8.008	141.2	37.62	9.490	2.717	0.861

[a] Not relevant for the present study.
[b] Not considered for the present study.
[c] Omitted for later study.

Table 4.3: $\ell\nu q\bar{q}$ cutflow for the six signal mass points.

Here it is seen that the reconstructed boson $p_\perp$ cuts, $p_\perp(\mathrm{J})$ and $p_\perp(\ell\nu$, completely kill the HVT(0.5 TeV), and to some extent HVT(1.0 TeV), mass point(s), while leaving the rest practically unchanged. The illustrates the need for separate (resolved) selections for the low resonance mass/boson $p_\perp$ signals, as was implement in the Run 1 analysis [16]. The HVT(0.5 TeV) mass point is included in the present study only to stress this point.

Below we list the impact of the major cuts in the event selection.

Cuts

The first major cut is the lepton veto, $N_{\mathrm{lepton}} = 1$. The distribution of the total number of veto leptons in the event is shown in Fig. 4.11. Here the white arrow indicates the (approximate) cut value.

Figure 4.11: Distribution of the total number of veto leptons in the events of the signal and background (stacked) samples considered, see App. A.3.
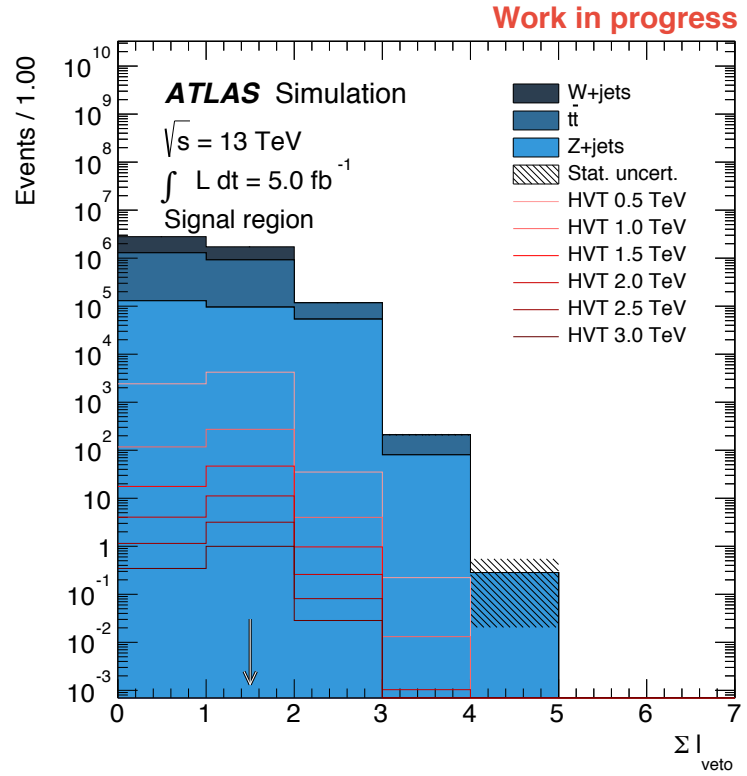
This is only approximately, since we require the presence of exactly one veto lepton, but this lepton should furthermore be selected as a signal lepton. Therefore, while the events in the bin marked by the arrow pass the necessary (but insufficient) condition $\sum \ell_{\mathrm{veto}} = 1$, a small fraction might fail the additional condition of $\sum \ell_{\mathrm{signal}} = 1$.

From this plot, we see that $Z + \mathrm{jets}$ events do indeed have more reconstructed leptons, as seen in the $\sum \ell_{\mathrm{veto}} = 2$ bin. $W + \mathrm{jets}$ and $t\bar{t}$ events contain only a true leptonic W, and therefore mainly have zero or one reconstructed lepton(s). The fact that so many zero-lepton events are reconstructed is partly due to the combined (for μ) and 'tight' (for e) identification requirements, see Sec. 1.4, along with several the cuts on several lepton isolation variables. For instance,

the electron 'tight' ID has an efficiency of 80% against electrons from truth W bosons in W + jets and $t\bar{t}$ events, while the same efficiency is $90 - 95\%$ for the various signal samples. The consequence of this, and similar cuts, can be seen in the $\sum \ell_{\text{veto}}$ distribution for the signal samples, where an excess at $\sum \ell_{\text{veto}} = 1$ is evident. This difference in efficiency, and similar efficiencies for the isolation variables, is in turn due to the fact that the leptons in the signal events generally have larger $p_{\perp}$ than those in the background events [44], which in itself is also contributing to killing electrons in soft background events, see Fig. 4.12.

Two of the other main cuts are the leading jet (J) and lepton plus neutrino ($\ell\nu$) $p_{\perp}$ cuts. For illustration, the leading jet $p_{\perp}$ distribution is shown in Fig. 4.13, where the arrow indicates the 400 GeV cut.
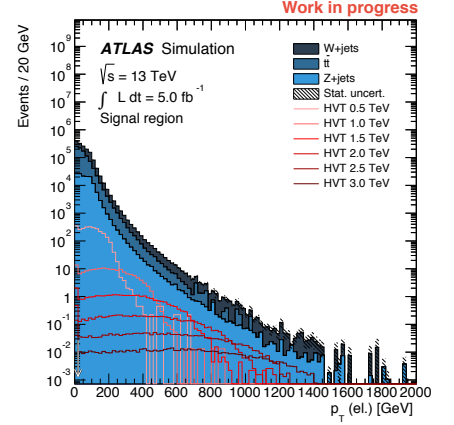
Figure 4.12: Distribution of electron $p_{\perp}$ in the signal and background (stacked) samples considered, see App. A.3.
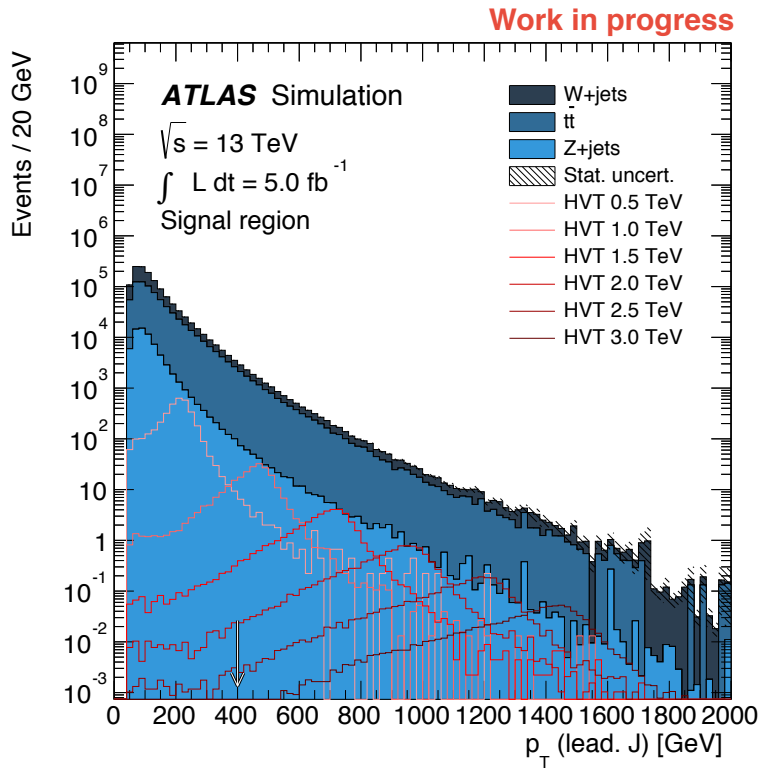
Figure 4.13: Distribution of leading $C/A^{R=1.2}$ jet $p_{\perp}$ in the signal and background (stacked) samples considered, see App. A.3.

We see the same non-resonant power-law structure present in the pure dijet-samples, see Figs. 3.1 and 4.2(a). It is also clear, that the choice to not include the lowest-$p_{\perp}$ W/Z + jets samples was reasonable, seeing that any effect at $50\,\text{GeV} < p_{\perp}(\text{J}) \lesssim 100\,\text{GeV}$, would be completely killed by the cut. The resonant structure of the signal $p_{\perp}$ spectra, seen in Fig. 4.1(a), is also recovered. Finally, we see that the $p_{\perp}(\text{J}) > 400\,\text{GeV}$ cut, ensuring the sufficient merging of the hadronic W decay products for clustering as a single fat jet cf. Eq. (1.53), all but kills the HVT(0.5 TeV) signal sample, as noted above.

The $p_{\perp}(\ell\nu)$ cut has a similar, but lesser, effect, due to the conservation of momentum in the transverse plane. The effect of the cut is

(quite naturally) seen to be largest for events which do not contain a true high-$p_\perp$ leptonic boson, with an efficiency of $\approx 15\%$ for $t\bar{t}$ and $Z+$jets, compared to 34.3% for $W+$jets and $\gtrsim 98\%$ for the high-mass signal samples.

The cuts on the leading jet mass and some substructure variable(s) (e.g. the momentum balance $\sqrt{y}$ for the last step in a mass-drop filtering procedure) are omitted in the initial event selection, since we want to study the impact of wavelet methods on the jet selection, and therefore we want to keep as many jets as possible to avoid any prior bias.

Finally, the b-jet veto is seen to discard mainly $t\bar{t}$ events (with an efficiency of 25%), while almost no other events (either $W/Z+$jets or signal), which is exactly the intended behaviour.

Overall, we see that the chosen cuts greatly remove the background, with overall efficiencies against $W+$jets, $t\bar{t}$, and $Z+$jets of 5.4×10^{-4}, 2.1×10^{-4}, and 1.5×10^{-4}, resp., while the high-mass signal samples generally have overall efficiencies of $\approx 60\%$.

■   Impact of wavelet analysis

We have now studied the effect of the various cuts imposed on the number of events—both signal and background—passing our selection. Similarly, we have looked in detail at a few of the more impactful cuts and their underlying distributions. The end result is a selection of events, which we believe to have a near-optimal composition of signal wrt. background events (in the absence of the jet mass and substructure cuts), and one which we can now attempt to improve on, using a wavelet-based analysis.

In the preceding sections, we found that a wavelet-based analysis using track information to perform a pile-up subtraction has the potential to no only improve the reconstructed jet mass resolution for boosted boson jets, but also to remove QCD jets around the boson pole mass(es), see e.g. Fig. 4.8. This could allow us to reject backgrounds with no true, hadronically decaying W bosons (i.e. the $W/Z+$jets components in Tab. 4.2, as well as QCD multijet events).

Therefore, after the above event selection, we create two collections: one of unmodified particle flow (PFlow) objects, and one of the same objects, but subjected to the $\mathrm{TF}(0.7) \oplus \mathrm{CL}(0.7)$ wavelet analysis with the track selection specified in Sec. 4.1. We then use the FASTJET package [77] to cluster jets with the $\mathrm{C/A}^{R=1.2}$ algorithm, using either of the two collections as inputs, and apply BDRS-A filter, cf. Sec. 1.4. Only jets (*now* groomed) with $p_\perp > 400\,\mathrm{GeV}$ are kept. In all events passing the selection above, the manually clustered jets in the two new jet collections are matched to the selected signal jet within[8] $dR < 0.6$. If no jets are matched, it is interpreted as the grooming procedure killing the signal jet in question, and the event is discarded. This would have occurred already in the event selection, had we chosen to apply the $p_\perp$ cut to the groomed jets, and such occurrences are thus expected. If, for a given collection, a jet is matched to the signal jet,

[8] This value was chosen based on the jet–jet-dR distribution in simulation.

this jet is flagged as the "new" signal jet within this collection and the event is kept.

For each collection, we now have a (slightly different) set of events with selected, groomed PFlow signal jets. Fig. 4.14 shows the distributions of the mass of the groomed signal jets for the unmodified and wavelet-cleaned event, resp.

(a) Unmodified.

(b) Wavelet-cleaned: $\text{TF}(0.7) \oplus \text{CL}(0.7)$.

Figure 4.14: Distributions of BDRS-A groomed jet masses, after event selection, for jets clustered in unmodified PFlow objects (left) and $\text{TF}(0.7) \oplus \text{CL}(0.7)$ wavelet-cleaned ones (right) for ATLAS MC signal and background samples, cf. App.A.3.
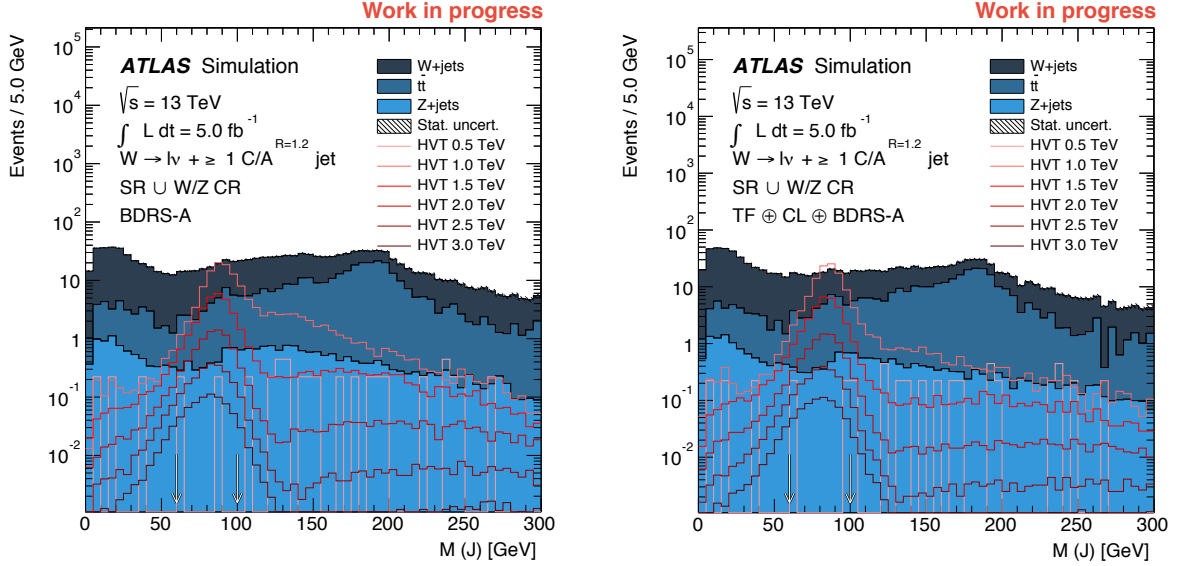
Here we see that the signal jets agree well with the W boson pole mass, as we already saw in Fig. 4.1(b), but that the background does not have a smoothly falling shape below $M_J \approx 200 \,\text{GeV}$, due to the $400 \,\text{GeV}$ $p_\perp$ cut. Even with the rather "brutal" logarithmic y-axis scale used presently, we see that, first, the signal jet mass peaks appear slightly more prominent over the background with wavelet-cleaning than without, and second, that the background spectra are shifted slightly downwards, see e.g. the $Z + \text{jets}$ "hump" at $M_J \approx 100 \,\text{GeV}$ and the $W + \text{jets}$ and $t\bar{t}$ humps at $M_J \approx 200 \,\text{GeV}$. This is an indication that we jet mass cut might benefit somewhat from the wavelet-based cleaning.

We can now impose the jet mass cut which was omitted in Tabs. 4.2 and 4.3, and, for each jet collection, require the signal jet to be within[9] $M_J \in [60, 100] \,\text{GeV}$, as indicated by the white arrows in Fig. 4.14. The resulting set of events then contains a jet which we believe resemble a hadronically decaying W boson. This, together with the reconstructed leptonically decaying W boson ($\ell\nu$), cf. Sec. 1.4, means that we can reconstruct the combined diboson system from the sum of their four-momenta. The invariant mass of this system should be a smoothly falling distribution for the background events, while we would expect peaks around the diboson resonance mass for each of the signal samples, cf. Sec. 1.1. The $M_{\ell\nu J}$ spectra are shown in Fig. 4.15 for unmodified and wavelet-cleaned events, resp.

[9] This cut defines the 'signal region' (SR). The events passing the original selection but failing the jet mass cut are customarily assigned to a '$W/Z + \text{jets}$ control region' (W/Z CR). Fig. 4.14 shows the union ($\cup$) of these two region, or equivalently, the distribution before the cut, in order to allow for a study of the full M_J range, and not just $[60, 100] \,\text{GeV}$.

(a) Unmodified.

(b) Wavelet-cleaned: $\mathrm{TF}(0.7) \oplus \mathrm{CL}(0.7)$.
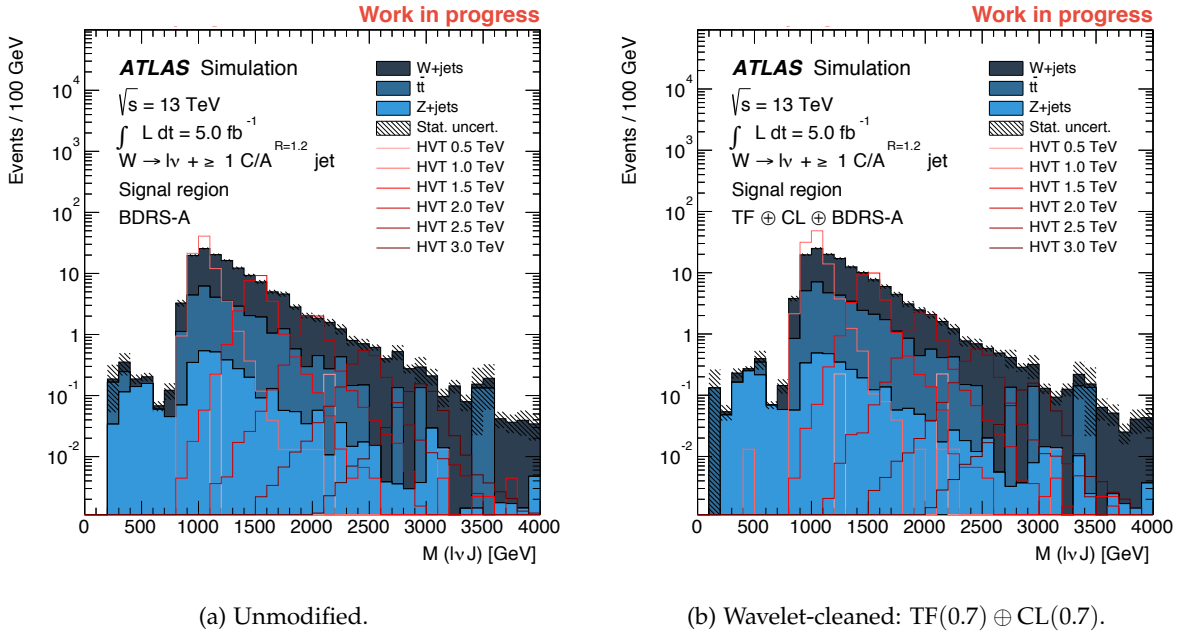
Figure 4.15: Distributions of reconstructed diboson masses $M_{\ell\nu J}$, after event selection and jet mass cut, using jets clustered in unmodified PFlow objects (left) and $\mathrm{TF}(0.7) \oplus \mathrm{CL}(0.7)$ wavelet-cleaned ones (right) for ATLAS MC signal and background samples, cf. App.A.3.

Here we see, as expected, a smoothly falling background distributions, truncated at $m_{\ell\nu J} \approx 800 - 1000\,\mathrm{GeV}$ since the reconstructed W bosons have $E \gtrsim 400\,\mathrm{GeV}$ and therefore the dibosons resonance, created roughly at rest, has a mass of $m_{WW} \approx 2E \gtrsim 800\,\mathrm{GeV}$.

Similarly to the case in Fig. 4.14, we see that the signal peaks in the wavelet-cleaned collection seem slightly more prominent, especially for the low-mass signal points. In order to quantify the improvement resulting from the wavelet-based cleaning, we compute the significance of each signal sample wrt. the background. Since the signal mass peaks in Fig. 4.15 are seen to match well with the generated resonance mass, see Tab. 4.1, we choose to perform a simple counting with in a $\pm 100\,\mathrm{GeV}$ range around each of the signal resonance masses, M_{HVT}. We will in general *not* have $S \ll B$, and therefore we use the combined signal significance $S/\sqrt{S+B}$ as our measure of sensitivity. The result is shown in Tab. 4.3

	M_{HVT} [TeV]	0.5	1.0	1.5	2.0	2.5	3.0
	$M_{\ell\nu J}$ range			$M_{\mathrm{HVT}} \pm 100\,\mathrm{GeV}$			
$\dfrac{S}{\sqrt{S+B}}$	Unmodified	0	5.99	2.91	1.43	0.65	0.31
	Wavelet-cleaned	0.02	7.15	3.15	1.42	0.67	0.32
	Improvement	–	19.3%	8.1%	−0.8%	2.8%	3.4%

Table 4.3: The diboson search sensitivity $S/\sqrt{S+B}$ using $\mathcal{L} = 5\,\mathrm{fb}^{-1}$ at $\sqrt{s} = 13\,\mathrm{TeV}$ with $\langle \mu \rangle \approx 25$, computed in intervals of $\pm 100\,\mathrm{GeV}$ around the generated mass for signal samples, resulting from groomed jets clustered in unmodified PFlow objects and PFlow objects cleaned using the $\mathrm{TF}(0.7) \oplus \mathrm{CL}(0.7)$ wavelet analysis.

From this table, we see that significant improvements in search sensitivity can be obtained using a wavelet-based cleaning, especially for moderate diboson resonance masses. In particular, we see that the improvements follow the same behaviour as seen e.g. in Fig. 4.9: for moderately low-$p_\perp$ jets, track-based wavelet methods have the potential to significantly improve the mass resolution for boosted boson jets, while simultaneously removing non-resonant QCD jets from the vicinity of $M_{W/Z}$. For increasing $p_\perp$, at fixed $\langle \mu \rangle$, this improvement diminishes, eventually yielding a performance which is equal to that of the baseline methods (jet grooming) within statistical errors. Therefore, the improvements shown in Tab. 4.3 for $M_{\mathrm{HVT}} \geq$ 2.0 TeV are consistent with 0 with errors.

As noted in Sec. 1.1, the diboson mass range of interest going into the LHC Run 2 is 0.8–1.5 TeV and above. This means that wavelet analyses have the potential to significantly increase the sensitivity in searches for new physics in the foreseeable future. Finally, we want to stress that the events used in this mock analysis have an average pile-up multiplicity of just $\langle \mu \rangle \approx 25$, which is already lower than the projections for the LHC Run 2. This means, that the improvements quoted in Tab. 4.3 are the *conservative* figures for a Run 2 analysis. Looking at Fig. 3.30, we may reasonably expect a similar pattern of improvement for higher $\langle \mu \rangle$. Therefore, even within a Run 2 context with $\langle \mu \rangle \sim 40$, we can expect improvements compared to what is shown in Tab. 4.3 by around a factor of two, *as well as* extension of the range of resonance masses, which experience an increase in sensitivity, also by a factor of two. For this reason, wavelet-based methods may have an immediate and substantial impact on searches for new resonances at the $\mathcal{O}(1-3\,\mathrm{TeV})$ scale.

Concluding remarks

You know what I'm craving?

A little perspective. That's it.

I'd like some fresh, clear,
well seasoned perspective.

Can you suggest a good wine to go with that?

— Anton Ego, *Ratatouille*

■ Summary

In this thesis, we have presented the mathematical tool of wavelets within the context of particle physics experiments and introduced a general prescription to apply wavelet-based analyses to hadron collision events. In addition, we have proposed and optimised a number of specific wavelet methods and studied their effect in detail, in both privately produced and official ATLAS Monte Carlo.

We found that wavelet-based methods, as a pile-up mitigation tool, have the potential to significantly improve measurements of both jet energies as well as jet masses, overall and on an per-event basis. In the case of jet mass resolution, a track-based wavelet analysis as been shown to yield improvements in performance beyond standard jet grooming techniques (when used in combination with these), and we have found that such an analysis not only improves the precision with which one is able to reconstruct the masses of hadronically decaying, high-$p_\perp$ vector bosons; it also actively "kills" non-resonant QCD jets, thus highlighting its usefulness for boson jet reconstruction. Specifically, we found that—for a range of $\langle \mu \rangle$ and jet $p_\perp$ values of interest for the LHC Run 2 and beyond—it is possible[10] to remove any jet energy measurement bias and increase the resolution by $\mathcal{O}(50\%)$ as well as increase the mass sensitivity for boosted boson jets with around $\mathcal{O}(100\%)$.

Finally, we have presented a mock analysis, using a search in the $\ell \nu J$ diboson channel as an example, in which we found that the sensitivity in such a search for narrow resonances beyond the Standard Model could be improved by upwards of 10% for resonance masses of $1 - 1.5\,\mathrm{TeV}$, even at relatively low pile-up multiplicities.

[10] Quoted values are for $\langle \mu \rangle = 40$ and $p_{\perp\,\mathrm{jet}} \approx 300\,\mathrm{GeV}$, using the metrics specified in Figs. 3.27 and 3.31.

■ Outlook

However, the present study is by no means complete. Rather, it represents a quick dash to the finish line, ignoring numerous interesting questions and research avenues along the way. For instance, this thesis has focused on reconstructing jet energies and masses, but any jet-based[11] observable might potentially benefit from wavelet analyses as a general tool. As an example, the reconstruction of missing transverse energy, heavily used in both exotics and SUSY searches, might very well benefit from a wavelet-based pile-up subtraction, especially under conditions of high luminosity.

[11] Possibly also any tracking- or calorimeter-based observable.

Similarly, we have presently only studied a small handful of potential wavelet methods, which are in no way guaranteed to be optimal. One might suspect that several new methods could be developed, ideally to suit the specific needs of various analysis tasks, building upon and improving beyond the ones presented here. Specifically, it might be possible to develop smarter, geometrically motivated methods, not relying on auxiliary information, e.g. through theoretical studies of the distributions of wavelet coefficient in minimum bias event or by using the RMS'es of these distributions as measured in data.

In the mock analysis, we studied the semi-leptonic channel, but as jet-cleaning tools, wavelet analyses may offer even greater sensitivity improvements in e.g. the fully hadronic diboson final state. Also, we omitted the question of jet substructure completely, which might given an incomplete picture: for the past five years, jet substructure techniques, for tagging hadronically decaying $W/Z/H/t$, have become popular, and are currently used most searches involving these particles. Since the wavelet analyses presented in this study remove particles from pile-up and scale the remainder to the level of the hard scatter, it is not unlikely that these will improve the separating capabilities of standard substructure methods beyond what is currently possible. What is more, at high $\langle \mu \rangle$ jet substructure methods may become unfeasible altogether, due to the flood of soft particles; but a particle-level pile-up mitigation method—like those proposed in this study—might change this prospect. In addition, it may be possible to use the wavelet coefficients, associated with a jet, *themselves* for tagging.

Notice also, that the methods presented here are not specific to ATLAS: the CMS and ALICE experiments, for instance, might in all likelihood benefit as just much, if not more, from techniques building of the foundation laid out in this thesis.

Finally, we want to stress that the positive results obtained in this thesis constitute a lower limit, showing only what is possible at a bare minimum. New methods may prove superior, as already hinted at, and the present results are only expected to improve with time, as the LHC—and the hadron colliders which will follow—proceed to operate under conditions of increasingly many simultaneous events. We therefore hope that the present study has done its part to indicate the promise of wavelet in a particle physics setting.

Appendices

A.1   Estimating neutrino pseudo-rapidity

We can write the four-momentum of a particle in our detector as

$$p^\mu = (E; p_x, p_y, p_z) = (M_\perp \cosh y; p_\perp \cos\phi, p_\perp \sin\phi, M_\perp \sinh y) \quad \text{(A.2)}$$

in the usual coordinate system, cf. Sec. 1.2.

Since the transverse mass is defined as

$$M_\perp^2 = E^2 - p_z^2 \tag{A.3}$$

we have, in the massless limit,

$$M_\perp^2 \xrightarrow[m=0]{} \mathbf{p}^2 - p_z^2 = p_\perp^2 \tag{A.4}$$

When dealing with a highly boosted vector boson decaying to leptons, which we in the selection in Sec. 4.2 require to have $p_\perp > 25\,\text{GeV}$, we can reasonably use the massless approximation $M_{\ell,\nu} \approx 0$ since $M_{\ell,\nu} \ll E_{\ell,\nu}$. Then the four-momenta of the decay products can we written in the simplified (purely geometrical) form

$$\ell^\mu = p_{\perp\ell} \left(\cosh\eta_\ell;\ \cos\phi_\ell,\ \sin\phi_\ell,\ \sinh\eta_\ell\right)$$
$$\nu^\mu = p_{\perp\nu} \left(\cosh\eta_\nu;\ \cos\phi_\nu,\ \sin\phi_\nu,\ \sinh\eta_\nu\right)$$

In these expressions we know all involved quantities from detector measurements, except the pseudo-rapidity of the neutrino, which cannot be reconstructed due to missing longitudinal momentum information escaping the detector's η coverage. (We identify the missing transverse energy and it's azimuthal angle with $p_{\perp\nu}$ and ϕ_ν, resp.) We then reconstruct the leptonically decaying W bosons as the four-momentum sum of the reconstructed lepton and neutrino, i.e.

$$W^\mu = \ell^\mu + \nu^\mu$$
$$= (p_{\perp\ell}\cosh\eta_\ell + p_{\perp\nu}\cosh\eta_\nu;\ p_{\perp\ell}\cos\phi_\ell + p_{\perp\nu}\cos\phi_\nu,$$
$$p_{\perp\ell}\sin\phi_\ell + p_{\perp\nu}\sin\phi_\nu,\ p_{\perp\ell}\sinh\eta_\ell + p_{\perp\nu}\sinh\eta_\nu) \tag{A.5}$$

and fixing the mass of the reconstructed W boson to the W boson pole mass [2], we obtain the relation (using two standard identities)

$$M_W^2 = W^2 = W^\mu W_\mu$$
$$= (p_{\perp\ell}\cosh\eta_\ell + p_{\perp\nu}\cosh\eta_\nu)^2 - (p_{\perp\ell}\cos\phi_\ell + p_{\perp\nu}\cos\phi_\nu)^2$$
$$- (p_{\perp\ell}\sin\phi_\ell + p_{\perp\nu}\sin\phi_\nu)^2 - (p_{\perp\ell}\sinh\eta_\ell + p_{\perp\nu}\sinh\eta_\nu)^2$$
$$= \left[(p_{\perp\ell})^2 \underbrace{\left(\cosh^2\eta_\ell - \sinh^2\eta_\ell\right)}_{1} + (p_{\perp\nu})^2 \underbrace{\left(\cosh^2\eta_\nu - \sinh^2\eta_\nu\right)}_{1} \right.$$
$$\left. + 2p_{\perp\ell}p_{\perp\nu}\left(\cosh\eta_\ell\cosh\eta_\nu - \sinh\eta_\ell\sinh\eta_\nu\right) \right]$$
$$- \left[(p_{\perp\ell})^2 \underbrace{\left(\cos^2\phi_\ell + \sin^2\phi_\ell\right)}_{1} + (p_{\perp\nu})^2 \underbrace{\left(\cos^2\phi_\nu + \sin^2\phi_\nu\right)}_{1} \right.$$
$$\left. + 2p_{\perp\ell}p_{\perp\nu}\left(\cos\phi_\ell\cos\phi_\nu + \sin\phi_\ell\sin\phi_\nu\right) \right]$$
$$= 2p_{\perp\ell}p_{\perp\nu}\Big[\ (\cosh\eta_\ell\cosh\eta_\nu - \sinh\eta_\ell\sinh\eta_\nu)$$
$$- (\cos\phi_\ell\cos\phi_\nu + \sin\phi_\ell\sin\phi_\nu) \Big] \tag{A.6}$$

and using two additional identities, as well as the knowledge that cosh and sinh are even and odd functions, resp.,

$$\cos(x - y) = \cos x \cos y + \sin x \sin y \qquad (A.7)$$

and

$$\begin{aligned}
\cosh(x + y) &= \cosh x \cosh y + \sinh x \sinh y \quad \Longrightarrow \\
\cosh(x - y) &= \cosh(x + (-y)) \\
&= \cosh x \cosh(-y) + \sinh x \sinh(-y) \\
&= \cosh x \cosh y - \sinh x \sinh y \qquad (A.8)
\end{aligned}$$

we obtain

$$M_W^2 = 2p_{\perp \ell} p_{\perp \nu} \left[\cosh(\eta_\ell - \eta_\nu) - \cos(\phi_\ell - \phi_\nu) \right] \qquad (A.9)$$

allowing us to isolate

$$\eta_\nu = \eta_\ell \pm \operatorname{arccosh} \left(\frac{M_W^2}{2p_{\perp \ell} p_{\perp \nu}} + \cos \Delta\phi \right) \qquad (A.10)$$

Since arccosh is defined on $[1, \infty[$, shown plotted in Fig. A.1, we have the following cases:

Argument is larger than one. Then the expression in Eq. (A.10) is well-defined and we have the choice between the two different solutions, given by the $\pm$ sign. Since both solutions (by construction) result in a reconstructed W boson on mass-shell, both solutions are equally viable in this respect. But since the true distribution of ν from W boson decay, differential in η, is maximal at $\nu = 0$ with tails out to larger values of $|\eta|$—cf. the discussion of the pseudo-rapidity plateau in Sec. 1.2—we choose the solution which gives the smallest $|\eta_\nu|$ value, simply because this is more likely to resemble the true situation.

Argument is smaller than one. Then the expression is ill-defined, meaning that there is no way to obtain a reconstructed W on mass-shell. However, since $\operatorname{arccosh} x \to 0$ as $x \to 1^+$, a reasonable assignment is to set the neutrino pseudo-rapidity equal to the signal lepton pseudo-rapidity. This will result in the lowest possible reconstructed W boson mass (and thus in the mass which is as close as possible to the pole mass) from Eq. (A.9).
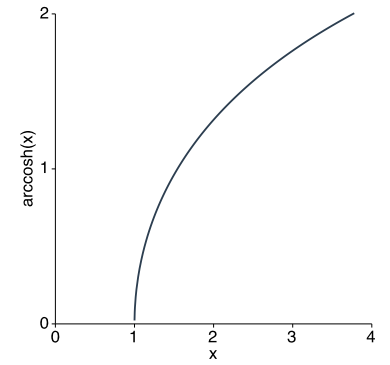
Figure A.1: Plot of arccosh x for a small range around $x = 1$.

A.2 Event displays

■ Events with $p_{\perp \mathrm{jet}} > 300\,\mathrm{GeV}$

Displays of the same hard scatter interaction ($W + $ jets) overlaid with two different average pile-up multiplicities, requiring $p_{\perp \mathrm{jet}} > 300\,\mathrm{GeV}$ and using the nominal wavelet setup, cf. Sec. 3.1. Jet circles extending outside the plot (at low and high ϕ-values) are to be understood as wrapping around the cylindrical detector.
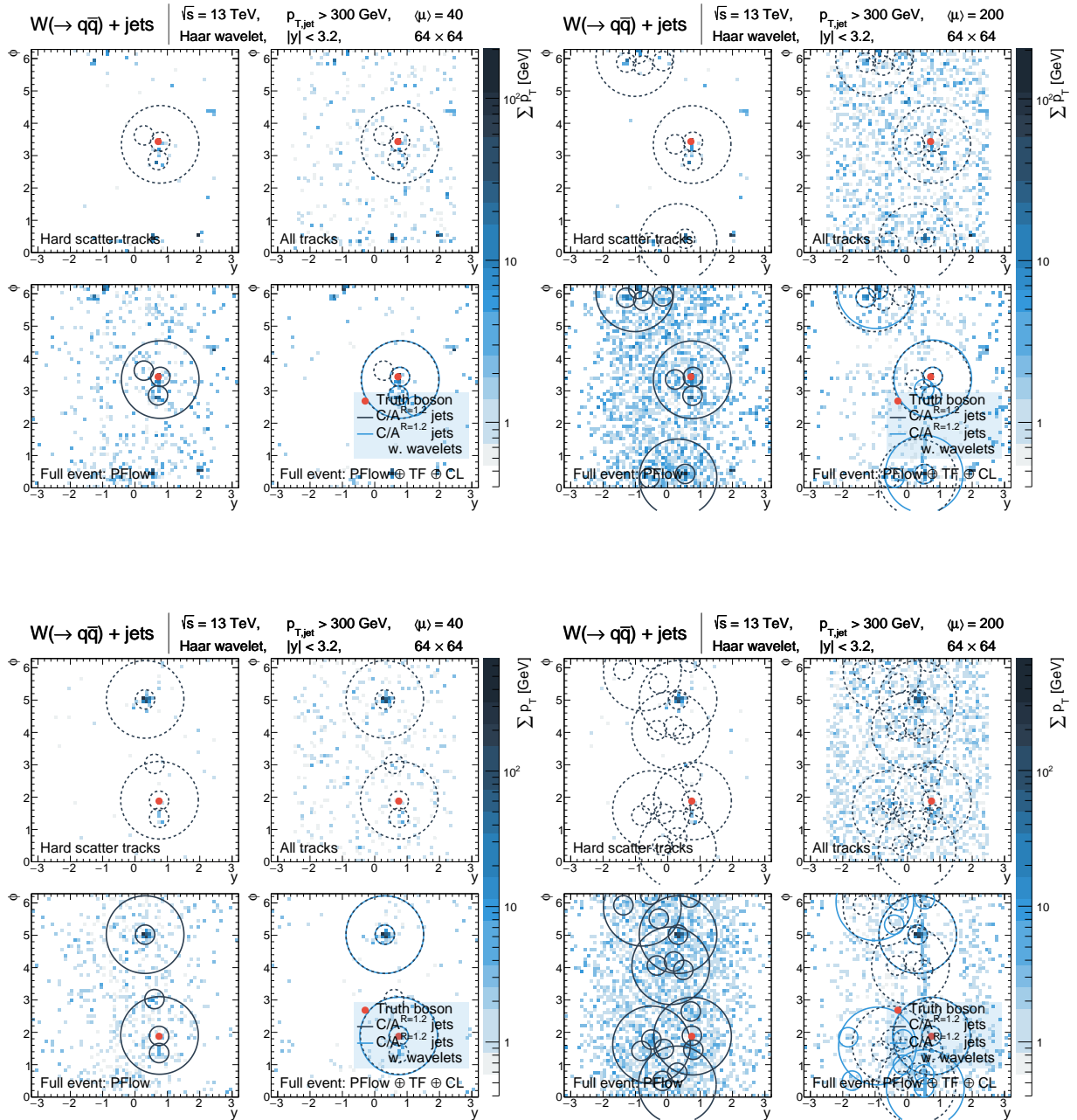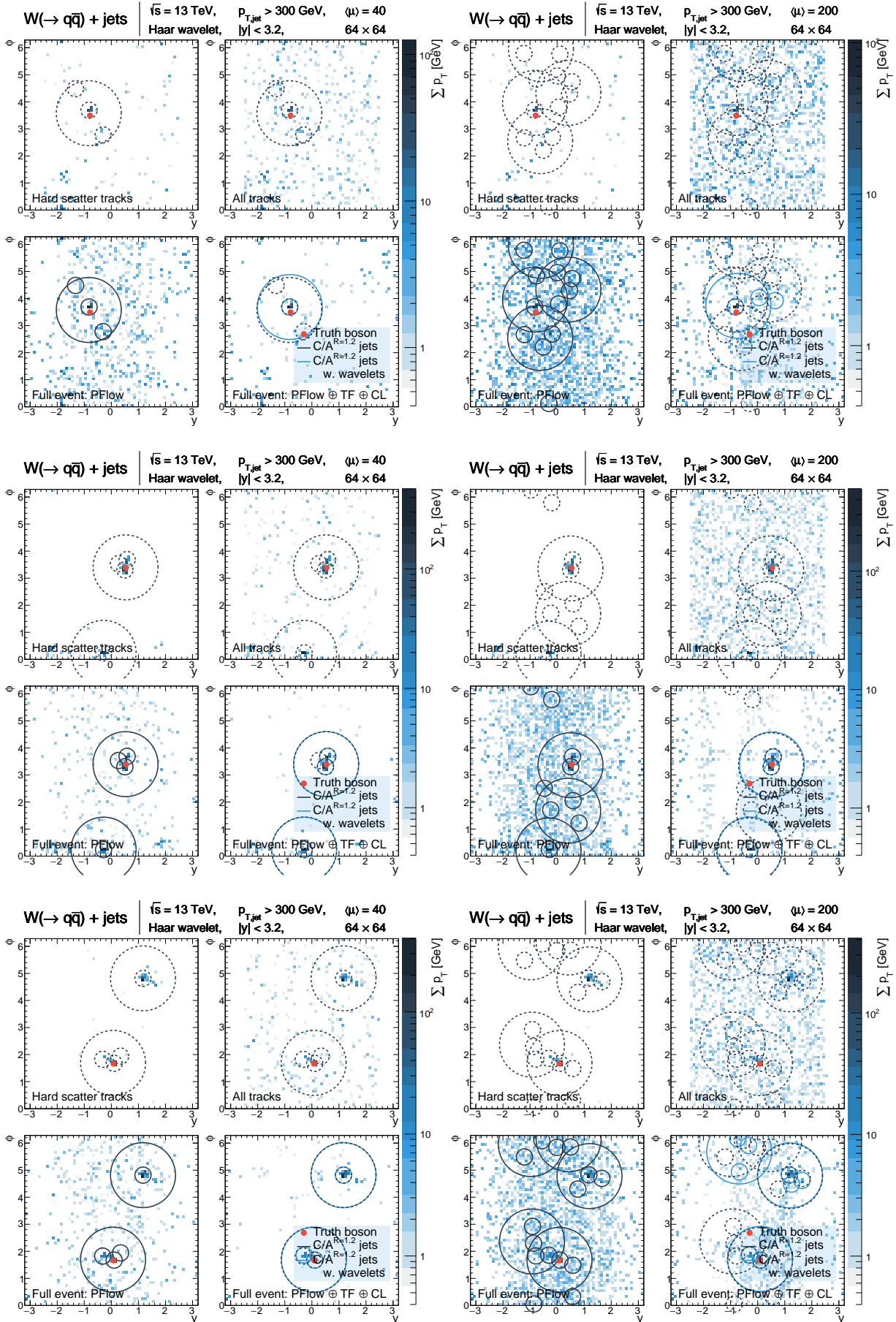
■ Events with $p_{\perp\text{jet}} > 1000\,\text{GeV}$

Displays of the same hard scatter interaction (W + jets) overlaid with two different average pile-up multiplicities, requiring $p_{\perp\text{jet}} > 1000\,\text{GeV}$ and using the nominal wavelet setup, cf. Sec. 3.1. Jet circles extending outside the plot (at low and high ϕ-values) are to be understood as wrapping around the cylindrical detector.
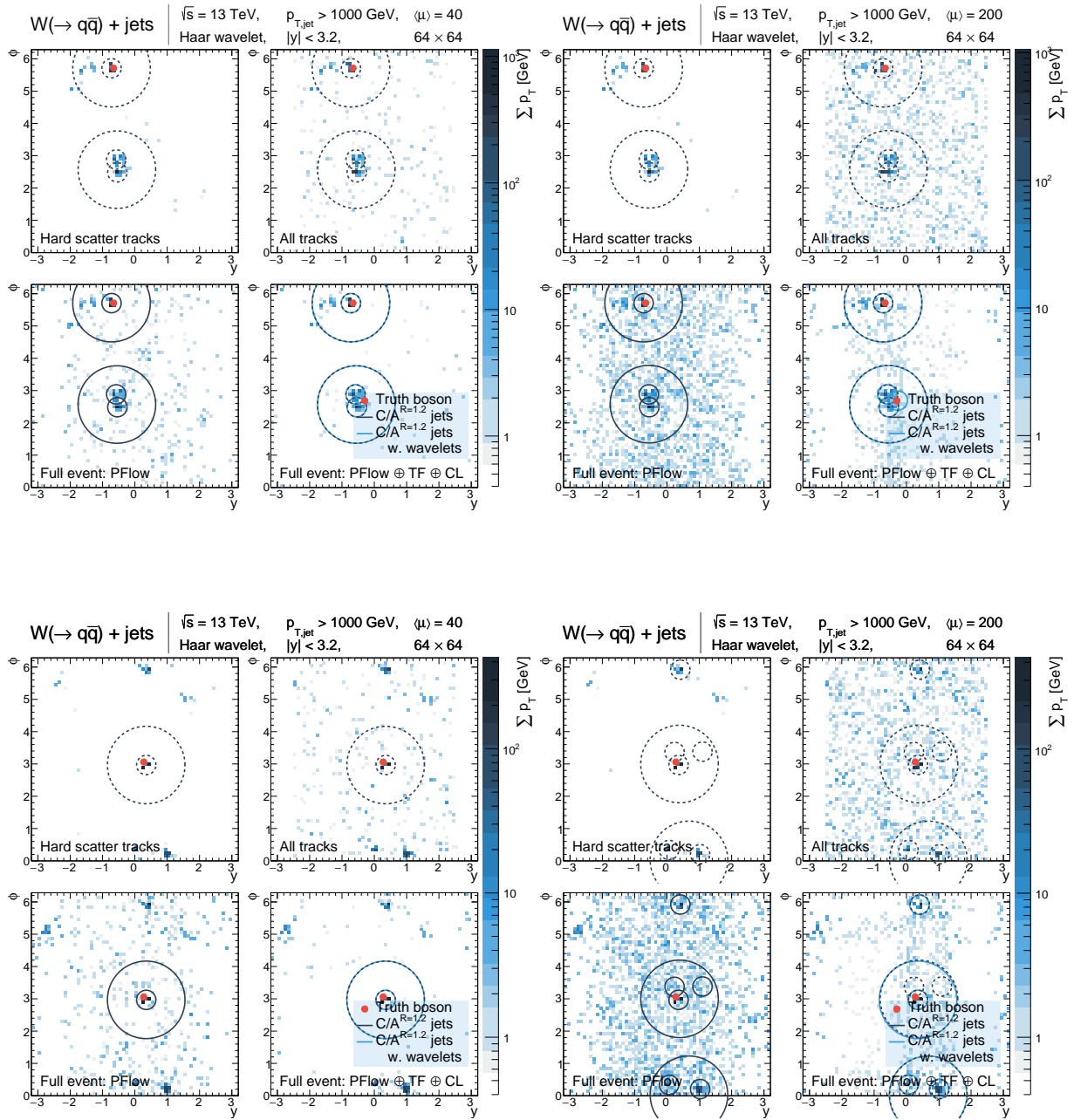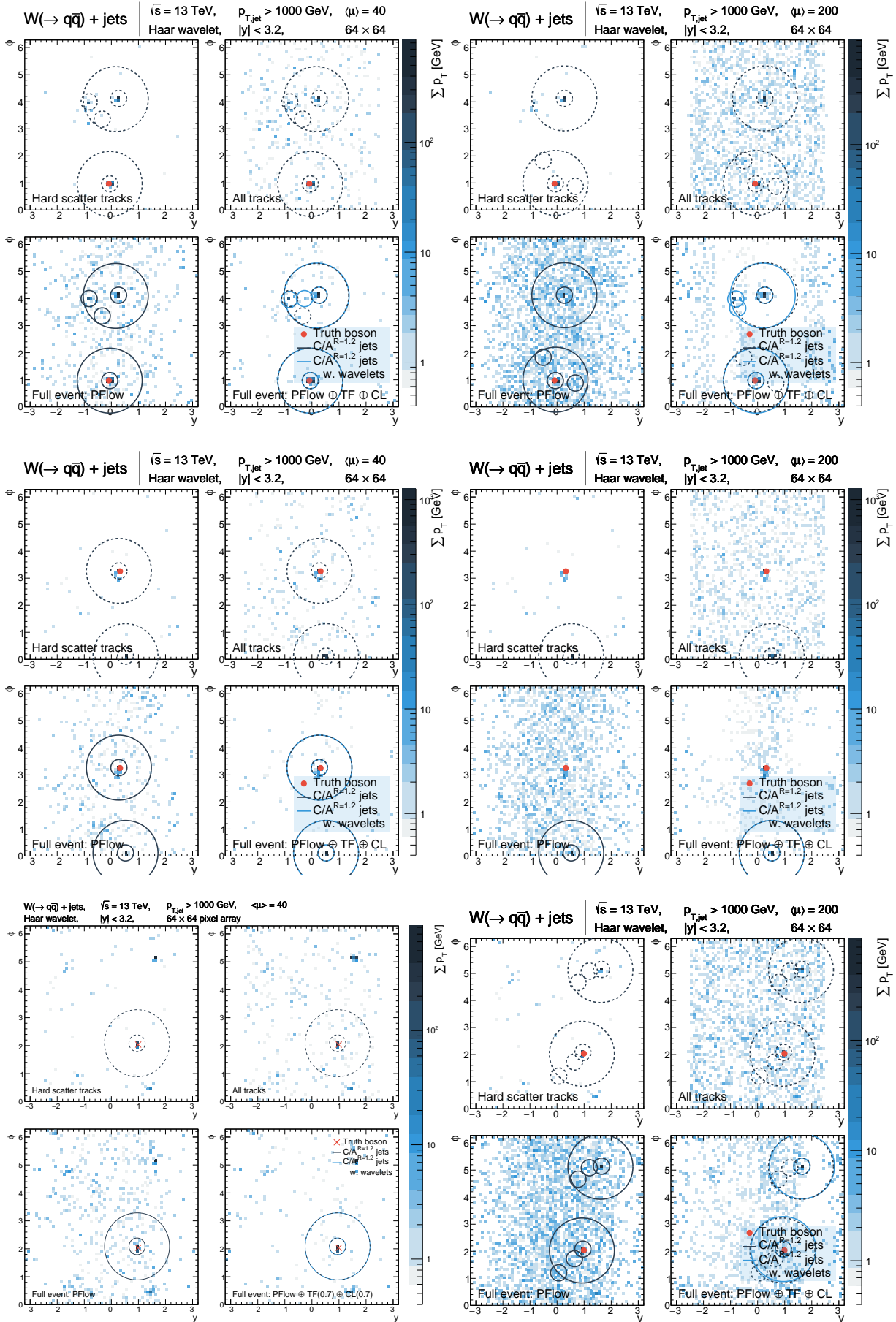
A.3   Diboson resonance search samples

■ Signal samples

The signal samples used in the simplified diboson resonance search in Sec. 4.2 comprise six different mass points for a heavy vector triplet (HVT) decaying semi-leptonically through two *W* bosons. All samples are generated under the mc14 simulation project at $\sqrt{s} = 13\,\text{TeV}$ and reconstructed using the ∗_r5787_r5853 reco. tag. These are the same samples used in studying the effectiveness of wavelet methods in ATLAS MC, cf. Tab. 4.1.

| Dataset ID	M_{HVT} [TeV]	Generator	$\sigma \times \text{BR}$ [fb]	$\varepsilon_{\text{filter}}$	$N_{\text{gen.}}$
Signal: HVT$(X\,\text{TeV}) \to \ell\nu q\bar{q}$					
203481	0.5		1765.	7.5616×10^{-1}	30000
203482	1.0		98.95	7.9607×10^{-1}	30000
203483	1.5		16.11	8.1147×10^{-1}	29500
203484	2.0	PYTHIA	3.81	8.1310×10^{-1}	30000
203485	2.5		1.076	8.1773×10^{-1}	30000
203486	3.0		0.3344	8.1999×10^{-1}	30000

Table A.1: HVT signal samples used in the simplified diboson resonance search in Sec. 4.2

■ Background samples

The background samples used in the simplified diboson resonance search in Sec. 4.2 represent the three of the major components found in [16], namely $W \to \ell\nu + X$, $t\bar{t} \to \ell + X$, and $Z \to \ell\ell + X$. All samples are generated under the mc14 simulation project at $\sqrt{s} = 13\,\text{TeV}$ and reconstructed using the ∗_r5787_r5853 reco. tag. The boson plus jets samples are generated in slices of $\hat{p}_\perp$, where we have chosen to use only samples with $\hat{p}_\perp > 70\,\text{GeV}$, since all samples with lower cuts will certainly be killed by the cuts in Sec. 4.2. The three samples within each boson plus jets decay mode (e.g. 167761, 167762, and 167763 for $W \to e\nu + X$ with $\hat{p}_\perp \in [70, 140]\,\text{GeV}$, see Tab. A.2) correspond to samples with generator *b*-jet filter (first line), *c*-jet filter and *b*-jet veto (second line), and *b*- and *c*-jet filter (third line).

Dataset ID	Decay mode	$p_\perp$ range [GeV]	Generator	$\sigma \times$ BR [fb]	$\varepsilon_{\text{filter}}$	$N_{\text{gen.}}$
$W(\to \ell\nu) + X$						
167761				557870.	5.4647×10^{-2}	349500
167762	$e\nu$			557920.	2.2664×10^{-1}	350000
167763				558370.	7.1828×10^{-1}	985500
167764				557870.	5.4620×10^{-2}	345500
167765	$\mu\nu$	[70, 140]		558250.	2.1980×10^{-1}	350000
167766				557480.	7.2404×10^{-1}	500000
167767				557920.	5.4733×10^{-2}	349500
167768	$\tau\nu$			557680.	2.2405×10^{-1}	349500
167769				558590.	7.1992×10^{-1}	999500
167770				81864.	7.4559×10^{-2}	198000
167771	$e\nu$			81918.	2.5481×10^{-1}	200000
167772				81764.	6.7089×10^{-1}	398000
167773				81775.	7.4477×10^{-2}	349500
167774	$\mu\nu$	[140, 280]		81813.	2.4981×10^{-1}	349000
167775				81925.	6.7640×10^{-1}	999500
167776				81867.	7.4564×10^{-2}	200000
167777	$\tau\nu$			81700.	2.5277×10^{-1}	199927
167778			SHERPA	81768.	6.7323×10^{-1}	400000
167779				6227.1	9.7603×10^{-2}	399500
167780	$e\nu$			6212.8	2.7254×10^{-1}	99500
167781				6235.7	6.2997×10^{-1}	200000
167782				6218.8	9.7362×10^{-2}	199500
167783[a]	$\mu\nu$	[280, 500]		6227.7	2.6644×10^{-1}	199000
167784[b]				6227.1	6.3520×10^{-1}	399500
167785				6223.5	9.7445×10^{-2}	399500
167786	$\tau\nu$			6231.1	2.7142×10^{-1}	100000
167787				6206.0	6.3116×10^{-1}	200000
167788				514.26	1.1791×10^{-1}	10000
167789	$e\nu$			520.84	2.8887×10^{-1}	10000
167790				513.06	5.9738×10^{-1}	40000
167791				514.91	1.1832×10^{-1}	398000
167792	$\mu\nu$	> 500		513.25	2.7951×10^{-1}	100000
167793				513.04	6.0274×10^{-1}	199000
167794				516.91	1.1872×10^{-1}	10000
167795	$\tau\nu$			513.62	2.9108×10^{-1}	10000
167796				514.46	5.9984×10^{-1}	40000

[a] Due to job submission errors, only 87% of available sample were processed. The remaining samples are weighted to take this into account.
[b] Due to job submission errors, only 98% of available sample were processed. The remaining samples are weighted to take this into account.

Table A.2: Dominant background samples used in the simplified diboson resonance search in Sec. 4.2: $W \to \ell\nu + X$.

Dataset ID	Decay mode	$p_\perp$ range [GeV]	Generator	$\sigma \times$ BR [fb]	$\varepsilon_{\text{filter}}$	$N_{\text{gen.}}$
$t\bar{t} \to \ell + X$						
110401[a]	–	–	PYTHIA	695840.	5.4316×10^{-1}	19963000
$Z(\to \ell\bar{\ell}) + X$						
167797				66749.	1.0172×10^{-1}	300000
167798	ee			66832.	3.9372×10^{-1}	99500
167799				66790.	5.0479×10^{-1}	100000
167800				66744.	1.0150×10^{-1}	299500
167801	$\mu\mu$	[70, 140]		66627.	3.9477×10^{-1}	100000
167802				66909.	5.0563×10^{-1}	99500
167803				66842.	1.0153×10^{-1}	299500
167804	$\tau\tau$			66882.	3.9342×10^{-1}	99500
167805				66995.	5.0584×10^{-1}	100000
167809[b]				10636.	1.1783×10^{-1}	200000
167810	ee			10621.	4.0744×10^{-1}	50000
167811				10617.	4.7278×10^{-1}	50000
167812				10629.	1.1818×10^{-1}	199500
167813	$\mu\mu$	[140, 280]		10650.	4.1070×10^{-1}	50000
167814				10675.	4.7554×10^{-1}	50000
167815				10626.	1.1824×10^{-1}	200000
167816	$\tau\tau$			10627.	4.0930×10^{-1}	50000
167817			SHERPA	10669.	4.7464×10^{-1}	50000
167821				830.57	1.3351×10^{-1}	99000
167822	ee			834.98	4.2293×10^{-1}	40000
167823				832.56	4.4542×10^{-1}	40000
167824				830.95	1.3286×10^{-1}	100000
167825	$\mu\mu$	[280, 500]		832.11	4.2498×10^{-1}	40000
167826				835.08	4.4470×10^{-1}	40000
167827				831.40	1.3267×10^{-1}	99500
167828	$\tau\tau$			833.35	4.2432×10^{-1}	40000
167829				830.13	4.4286×10^{-1}	40000
167833				68.37	1.4561×10^{-1}	9500
167834	ee			68.45	4.3408×10^{-1}	10000
167835				68.50	4.1906×10^{-1}	40000
167836				68.31	1.4377×10^{-1}	10000
167837	$\mu\mu$	> 500		69.01	4.4096×10^{-1}	10000
167838				68.74	4.1689×10^{-1}	40000
167839				68.46	1.4378×10^{-1}	10000
167840	$\tau\tau$			68.75	4.4112×10^{-1}	10000
167841				67.865	4.1496×10^{-1}	40000

[a] Due to grid job submission errors, only 72.1% of available sample were processed. The remaining samples are weighted to take this into account.

[a] Due to grid job submission errors, only 95% of available sample were processed. The remaining samples are weighted to take this into account.

Table A.3: Sub-dominant background samples used in the simplified diboson resonance search in Sec. 4.2: $t\bar{t} \to \ell + X$ and $Z \to \ell\ell + X$.

References

[1] L. H. RYDER. *Quantum Field Theory*. Cambridge University Press, 2nd edition, 1996.

[2] PARTICLE DATA GROUP (K. A. OLIVE *et al.*). Review of Particle Physics. *Chin. Phys. C*, **8**:090001, 2014.

[3] ATLAS AND CMS COLLABORATIONS (G. AAD *et al.*). Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and $8\,\text{TeV}$ with the ATLAS and CMS Experiments. *Phys. Rev. Lett.*, **114**:191803, 2015. [arXiv:1503.07589].

[4] D. GREEN. *High P_T Physics at Hadron Colliders*. Cambridge Monographs on Particle Physics, Nuclear Physics, and Cosmology 22. Cambridge University Press, 2005.

[5] B. R. MARTIN AND G. SHAW. *Particle Physics*. Manchester Physics series. John Wiley and Sons, Ltd, 3rd edition, 2008.

[6] R. BROUT AND F. ENGLERT. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.*, **13**(9):321–323, 1964.

[7] P. HIGGS. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, **13**(16):508–509, 1964.

[8] G. GURALNIK, C. R. HAGEN, AND T. W. B. KIBBLE. Global Conservation Laws and Massless Particles. *Phys. Rev. Lett.*, **13**(20):585–587, 1964.

[9] ATLAS COLLABORATION (G. AAD *et al.*). Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B.*, **716**:1–29, 2012. [CDS:1124337].

[10] CMS COLLABORATION (S. CHATRCHYAN *et al.*). Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B.*, **716**:30–31, 2012. [CDS:1124338].

[11] P. H. DAMGAARD *et al.* Constraints on New Physics from Baryogenesis and Large Hadron Collider Data. *Phys. Rev. Lett.*, **111**:221804, 2013. [arXiv:1305.4362].

[12] L. RANDALL AND S. SUNDRUM. A Large Mass Hierarchy from a Small Extra Dimension. *Phys. Rev. Lett.*, **83**:3370–3373, 1999. [arXiv:hep-ph/9905221].

[13] G. ALTARELLI, B. MELE, AND M. RUIZ-ALTABA. Searching for New Heavy Vector Bosons in $p\bar{p}$ Colliders. *Z. Phys. C*, **45**:109, 1989. [inSPIRE:279185].

[14] ATLAS COLLABORATION. Search for high-mass diboson resonances with boson-tagged jets in proton?proton collisions at $\sqrt{s} = 8\,\text{TeV}$ with the ATLAS detector. [arXiv:1506.00962], 2015.

[15] ATLAS COLLABORATION. Search for resonant diboson production in the $\ell\ell q\bar{q}$ final state in pp collisions at $\sqrt{s} = 8\,\text{TeV}$ with the ATLAS detector. *Eur. Phys. J. C*, **75**(69), 2015. [arXiv:1409.6190].

[16] ATLAS COLLABORATION. Search for production of WW/WZ resonances decaying to a lepton, neutrino and jets in pp collisions at $\sqrt{s} = 8\,\text{TeV}$ with the ATLAS detector. *Eur. Phys. J. C*, **75**(209), 2015. [arXiv:1503.04677].

[17] NNPDF (R. D. BALL *et al.*). Impact of Heavy Quark Masses on Parton Distributions and LHC Phenomenology. *Nucl. Phys. B*, **849**:296–363, 2013. [arXiv:1207.1303].

[18] P. SKANDS. Introduction to QCD. Lecture notes for TASI 2012. [arXiv:1207.2389].

[19] J. M. CAMPBELL, J. W. HUSTON, AND W. J. STIRLING. Hard Interactions of Quarks and Gluons: A Primer for LHC Physics. *Rept. Prog. Phys.*, **70**:89, 2007. [arXiv:hep-ph/0611148].

[20] A. BUCKLEY *et al.* General-purpose event generators for LHC physics. *Phys. Rept.*, **504**(504):145–233, 2011. [inSPIRE:884202].

[21] M. SREDNICKI. *Quantum Field Theory*. Cambridge University Press, 2007.

[22] T. GLEISBERG *et al.* Event generation with SHERPA 1.1. *JHEP*, **02**:007, 2009. [arXiv:0811.4622].

[23] T. SJÖSTRAND, S. MRENNA, AND P. SKANDS. PYTHIA 6.4 Physics and Manual. *JHEP*, **05**:026, 2006. [arXiv:hep-ph/0603175].

[24] M. BAHR *et al.* Herwig++ Physics and Manual. *Eur. Phys. J. C*, **58**:639–707, 2008. [arXiv:0803.0883].

[25] J. M. CAMPBELL, W. T. GIELE, AND C. WILLIAMS. The Matrix Element Method at Next-to-Leading Order. *JHEP*, **11**:043, 2012. [arXiv:1204.4424].

[26] G. Altarelli and G. Parisi. Asymptotic Freedom in Parton Language. *Nucl. Phys. B*, **128**:298–318, 1977. [inSPIRE:119585].

[27] T. Sjöstrand. Particle Physics Phenomenology. Lecture slides, Lund University, 2015.

[28] ATLAS Collaboration. Charged-particle multiplicities in *pp* interactions measured with the ATLAS detector at the LHC. *New J. Phys.*, **13**(053033), 2011. [arXiv:1012.5104].

[29] L. Evans and P. Bryant (eds.). LHC Machine. *JINST*, **3**:S08001, 2008. [inSPIRE:796247].

[30] P. Hansen. Particle detectors and accelerators. Lecture notes, University of Copenhagen, 2015, 1st ed.

[31] ATLAS Collaboration (G. Aad *et al.*). Measurement of the total cross section from elastic scattering in *pp* collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *Nucl. Phys. B*, **889**:486–548, 2014. [arXiv:1408.5778].

[32] ATLAS Collaboration (G. Aad *et al.*). The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, **3**(S08003), 2008. [inSPIRE:796888].

[33] CMS Collaboration (S. Chatrchyan *et al.*). The CMS Experiment at the CERN LHC. *JINST*, **3**(S08004), 2008. [inSPIRE:796887].

[34] ATLAS Collaboration (G. Aad *et al.*). ATLAS pixel detector electronics and sensors. *JINST*, **7**(P07007), 2008. [inSPIRE:809652].

[35] ATLAS IBL Community for the ATLAS Collaboration (G. Aad *et al.*). *ATLAS Insertable B-Layer Technical Design Report*. ATLAS TDR 19, CERN/LHCC 2010-013, 2010. [CDS:1291633].

[36] S. Tavernier. *Experimental Techniques in Nuclear and Particles Physics*. Springer, 2010.

[37] T. Cornelissen *et al.* Concepts, Design and Implementation of the ATLAS New Tracking (NEWT). Technical Report ATL-SOFT-PUB-2007-007. ATL-COM-SOFT-2007-002, CERN, Geneva, Mar 2007. [CDS:1020106].

[38] ATLAS Collaboration. A neural network clustering algorithm for the ATLAS silicon pixel detector. *J. Instrum.*, **9**:P09009, 2014. [arXiv:1406.7690].

[39] R. Mankel. *Rep. Prog. Phys.*, **64**:553, 2004. [arXiv:physics/0402039].

[40] Performance of primary vertex reconstruction in proton-proton collisions at $\sqrt{s} = 7$ TeV in the ATLAS experiment. Technical Report ATLAS-CONF-2010-069, CERN, Geneva, Jul 2010.

[41] W. Lampl *et al.* Calorimeter Clustering Algorithms: Description and Performance. ATLAS NOTE: ATL-LARG-PUB-2008-002 `[CDS:1099735]`, 2008.

[42] M. Hodgkinson, D. Tovey, and R. Duxfield. Energy Flow Reconstruction with the eflowRec Combined Reconstruction Software in Athena 15.6.9.8. Technical Report ATL-PHYS-INT-2011-031, CERN, Geneva, Apr 2011. `[CDS:1342125]`. Can be made available upon request.

[43] Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. Technical Report CMS-PAS-PFT-09-001, CERN, Geneva, Apr 2009. `[CDS:1194487]`.

[44] ATLAS Collaboration (G. Aad *et al.*). Electron reconstruction and identification efficiency measurements with the atlas detector using the 2011 lhc proton-proton collision data. *Eur. Phys. J*, **C74**:2941, 2014. `[arXiv:1404.2240]`.

[45] ATLAS Collaboration (G. Aad *et al.*). Expected performance of the atlas experiment - detector, trigger and physics. `[arXiv:0901.0512]`, 2009.

[46] G. Salam. IPMU-YITP School and Workshop on Monte Carlo Tools for LHC. Lecture slides, 2011. `[http://conference.ippp.dur.ac.uk/event/309/material/slides/0?contribId=4]`.

[47] G. Salam. Towards Jetography. *Eur. Phys. J.*, **C67**:637–686, 2010. `[arXiv:0906.1833]`.

[48] Yu. L. Dorkshitzer *et al.* Longitudinally invariant K_t clustering algorithms for hadron hadron collisions. *Nucl. Phys.*, **B406**:187–224, 1993. `[CDS:354686]`.

[49] Yu. L. Dorkshitzer *et al.* Better Jet Clustering Algorithms. *JHEP*, **08**:001, 1997. `[arXiv:hep-ph/9707323]`.

[50] M. Cacciari, G. P. Salam, G. Soyez. Better Jet Clustering Algorithms. *JHEP*, **04**:063, 2008. `[arXiv:0802.1189]`.

[51] J. M. Butterworth *et al.* Jet substructure as a new Higgs search channel at the LHC. *Phys. Rev. Lett.*, **100**:242001, 2008. `[arXiv:0802.2470]`.

[52] ATLAS Collaboration. Performance of jet substructure techniques for large-R jets in proton-proton collisions at sqrt(s) = 7 TeV using the ATLAS detector. *JHEP*, **09**:076, 2013. `[arXiv:1306.4945]`.

[53] D. Krohn, J. Thaler, and L.-T.Wang. Jet Trimming. *JHEP*, **02**:084, 2010. `[arXiv:0912.1342]`.

[54] S. D. Ellis, C. K. Vermillion, and J. R. Walsh. Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches. *Phys. Rev. D*, **81**:094023, 2010. `[arXiv:0912.0033]`.

[55] Pile-up subtraction and suppression for jets in ATLAS. Technical Report ATLAS-CONF-2013-083, CERN, Geneva, Aug 2013. `[CDS:1570994]`.

[56] M. Cacciari, G. P. Salam. Pileup subtraction using jet areas. *Phys. Lett. B*, **659**:119–126, 2008. `[arXiv:0707.1378]`.

[57] M. Cacciari, G. P. Salam, and G. Soyez. The Catchment Area of Jets. *JHEP*, **04**:005, 2008. `[arXiv:0802.1188]`.

[58] M. Cacciari, G. P. Salam, and G. Soyez. On the use of charged-track information to subtract neutral pileup. *Phys. Rev. D*, **92**:014003, 2015. `[arXiv:1404.7353]`.

[59] M. Cacciari, G. P. Salam, and G. Soyez. SoftKiller, a particle-level pileup removal method. *Eur. Phys. J. C.*, **75**(2):59, 2015. `[arXiv:1407.0408]`.

[60] D. Betrolini *et al.* Pileup Per Particle Identification. *JHEP*, **10**:59, 2014. `[arXiv:1407.6013]`.

[61] I. Daubechies. *Ten Lectures on Wavelets*. CBMS-NDF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), 1992.

[62] A. Jensen and A. la Cour-Harbo. *Ripples in Mathematics: the Discrete Wavelet Transform*. Springer, 2001.

[63] J. Williams and K. Amaratunga. Introduction to Wavelets in Engineering. *Int. Journ. Num. Meth. Eng.*, **37**(14):2365–2388, 1994.

[64] A. Haar. Zur Theorie der orthogonalen Funktionensysteme. *Mathematische Annalen*, **69**(3):331–371, 1910.

[65] M. Galassi *et al.* GNU Scientific Library Reference Manual (3rd Ed.). ISBN 0954612078, `http://www.gnu.org/software/gsl/`.

[66] L. Tenorio *et al.* Applications of wavelets to the analysis of cosmic microwave background maps. *Mon. Not. R. Astron. Soc.*, **310**:823–834, 1999. `[arXiv:astro-ph/9903206]`.

[67] J. W. Monk. Wavelet Analysis: Event De-noising, Shower Evolution and Jet Substructure Without Jets. `[arXiv:1405.5008]`, `http://newwave.hepforge.org/`, 2014.

[68] ATLAS Collaboration. Jet energy resolution in proton-proton collisions at $\sqrt{s} = 7\,\mathrm{TeV}$ recorded in 2010 with the ATLAS detector. *Eur. Phys. J. C*, **73**(2306), 2013. `[arXiv:1210.6210]`.

[69] G. Miu and T. Sjöstrand. W Production in an Improved Parton-Shower Approach. `[arXiv:hep-ph/9812455]`, 1998.

[70] T. Sjöstrand *et al.* An Introduction to pythia 8.2. `[arXiv:1410.3012]`, 2014.

[71] ATLAS Collaboration (A. Buckley *et al.*). ATLAS Pythia 8 tunes to 7 TeV data. ATLAS NOTE: ATL-PHYS-PUB-2014-021 [CDS:1966419], 2014.

[72] NNPDF (R. D. Ball *et al.*). Parton distributions with LHC data. *Nucl. Phys. B*, **867**:244–289, 2013. [arXiv:1207.1303].

[73] C. Roehr *et al.* http://pilemc.hepforge.org/.

[74] ATLAS Collaboration (A. Buckley *et al.*). Summary of AT-LAS Pythia 8 tunes. ATLAS NOTE: ATL-PHYS-PUB-2012-003 [CDS:1474107], 2012.

[75] A. Buckley *et al.* Rivet user manual. [arXiv:1003.0694], http://rivet.hepforge.org/, 2010.

[76] R. Brun and F. Rademakers. ROOT - An Object Oriented Data Analysis Framework. *Nucl. Inst. & Meth. in Phys. Res. A*, **389**:81–86, 1997. See also http://root.cern.ch/.

[77] M. Macciari, G. P. Salam, and G. Soyez. FastJet user manual. *Eur. Phys. J. C*, **72**(1896), 2012. [arXiv:1111.6097], http://www.fastjet.fr/.

[78] Pile-up corrections for jets from proton-proton collisions at $\sqrt{s} =$ 7 TeV in ATLAS in 2011. Technical Report ATLAS-CONF-2012-064, CERN, Geneva, Jul 2012. [CDS:1459529].

[79] Charged-particle distributions in sqrt(s)=13 TeV pp interactions measured with the ATLAS detector at the LHC. Technical Report ATLAS-CONF-2015-028, CERN, Geneva, Jul 2015. [CDS:2037701].

[80] MSTW (A. D. Martin *et al.*). Parton distributions for the LHC. *Eur. Phys. J. C*, **63**:189–285, 2009. [arXiv:0901.0002].

[81] J. Alwall *et al.* The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, **07**:079, 2014. [arXiv:1405.0301].

[82] CTEQ (J. Pumplin *et al.*). New generation of parton distributions with uncertainties from global QCD analysis. *JHEP*, **07**:012, 2002. [arXiv:hep-ph/0201195].

[83] O. Harris and G. Watts. *Search for Weakly-interacting Long-lived Particles in Proton-Proton Collisions at sqrts = 8 TeV with the ATLAS Detector*. PhD thesis, Washington U., Seattle, 2013. [CDS:1601599], Presented 2013.

[84] S. Agostinelli *et al.* Geant4 - a simulation toolkit. *Nucl. Instr. Methods Phys. Res. A*, **506**:250–303, 2003.

[85] CTEQ (H.-L. Lai *et al.*). New parton distributions for collider physics. *Phys. Rev. D*, **82**:074024, 2010. [arXiv:1007.2241].

[86] P. Nason. A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms. *JHEP*, **11**:040, 2004. [arXiv:hep-ph/0409146].

[87] P. Skands. Tuning Monte Carlo Generators: The Perugia Tunes. *Phys. Rev. D*, **82**:074018, 2010. [arXiv:1005.3457].

[88] ATLAS Collaboration (T. Barillari *et al.*). Local Hadronic Calibration. Technical Report ATL-LARG-PUB-2009-001-2, CERN, Geneva, Jun 2008. [CDS:1112035].

[89] ATLAS Collaboration (G. Aad *et al*). A measurement of the ratio of the production cross sections for W and Z bosons in association with jets with the ATLAS detector. *Eur. Phys. J. C*, **74**:3168, 2014. [arXiv:1408.6510].

BOOSTED BOSONS
AND
WAVELETS