direction of the secondary vertex relative to the primary vertex. Six of these distributions, evaluated on the $t\bar{t}$ simulated sample, are illustrated in Fig. 3.14.

JetFitter, the topological multi-vertex finding algorithm: As opposed to SV1, which assumes the presence of just one SV in heavy-flavour jets, JetFitter [73] aims at simultaneously reconstructing the full b-c hadron decay chain, assuming that the decay vertices lie on the line defined by the trajectory of the b-hadron, initialised as the jet axis. The assumption implies that all the particles emerging from the decay of the heavy-flavour hadrons must originate from a point on such line. This constraint allows the reconstruction of incomplete topologies in which only a single track from the decay of the heavy-flavour hadron is selected.

A track selection is designed to minimize the tracks not originating from the heavy-flavour hadrons decay. Tracks compatible with the PV are removed. Then, track pairs are fitted to identify two-track vertices. Tracks forming two-track vertices located in regions occupied by the pixel layers are discarded as they likely emerge from interactions with the material. Tracks forming vertices compatible with photon conversions or K_S^0 and Λ^0 decays are also discarded.

An iterative process of vertex finding and fitting is used to identify the most probable number of vertices in the decay chain and fit their positions and the b-hadron trajectory. Initially, all tracks are considered as distinct one-track vertices. A modified Kalman filter [74] fit is applied, extracting information about the full decay chain, encoded in the position of the PV, the flight direction of the b-hadron, and the distances of the secondary vertices to the PV. Vertices with a χ^2 contribution to the overall fit below a 0.1% probability are removed.

Next, a clustering step evaluates all pairs of vertices to determine the probability that each pair belongs to a common vertex and that the vertex is part of the decay chain. This is computed using the same fitting technique. The two vertices with the highest probability are merged, and the process repeats with the new decay chain structure. The iterations continue until all pairs of vertices above a certain vertex mass-dependent threshold have been combined.

In the final stage, tracks that did not form any two-track vertices in the initial iteration are evaluated if they meet strict p_T and IP criteria. The fit procedure is reapplied to see if these tracks can form additional single-track vertices along the b-hadron trajectory.

The algorithm's output is a topology where tracks are uniquely associated with vertices along the path of the b-hadron. Two sets of discriminant variables are extracted to be inputs for high-level taggers. The first set, including the track multiplicity at the JetFitter displaced vertices, the invariant mass of tracks associated with these vertices, their energy fraction and their average three-dimensional decay length significance, is evaluated on all the jets. Some of these variables are displayed in Fig. 3.15. The second set is related only to jets in which a single secondary vertex is reconstructed with an intermediate charged
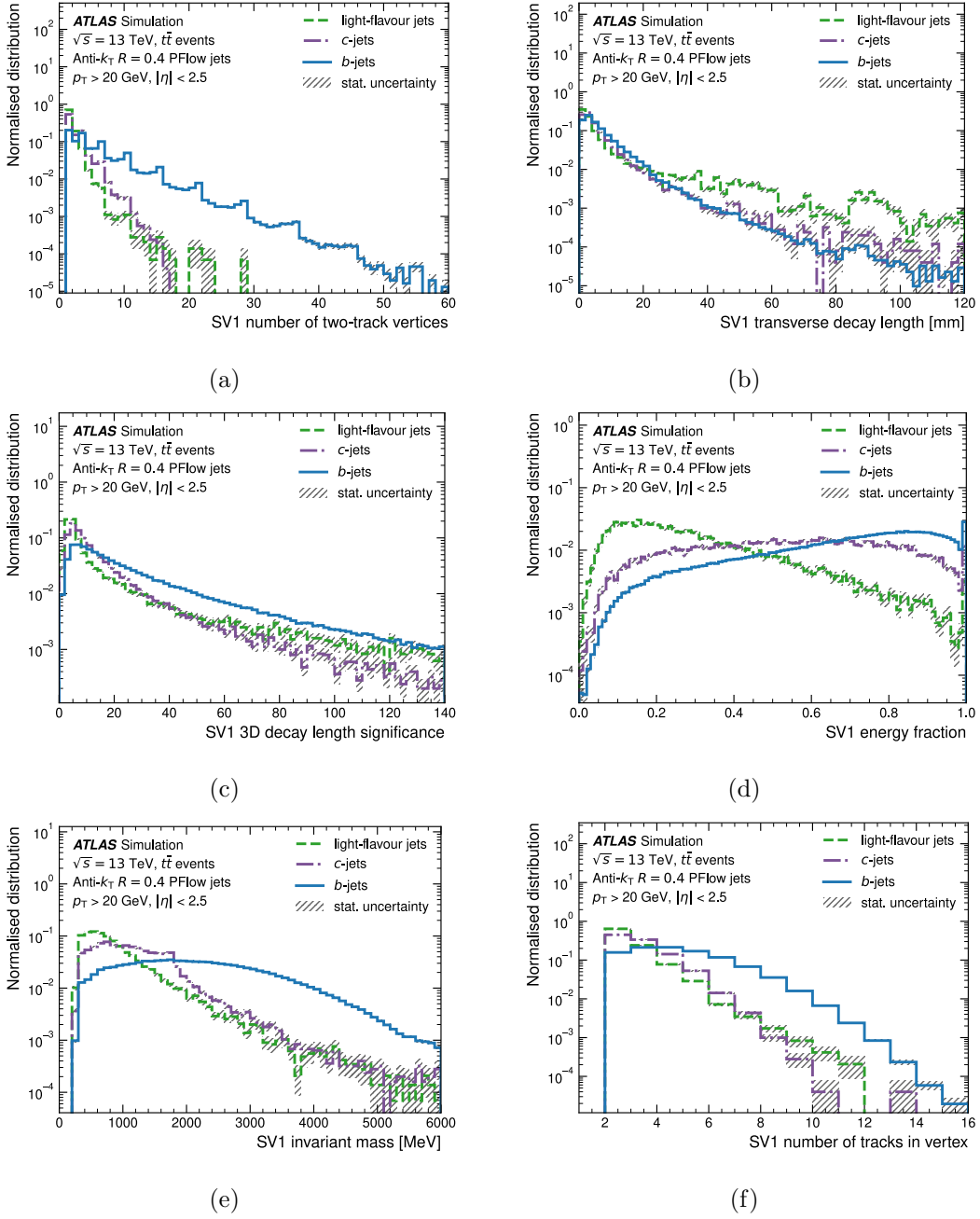
(a)

(b)

(c)

(d)

(e)

(f)

Figure 3.14: Properties of secondary vertices reconstructed by the `SV1` algorithm for b-jets, c-jets and light-flavour jets in $t\bar{t}$ simulated events: (a) the number of two-track vertices reconstructed within the jet, (b) the transverse decay length, (c) the 3D decay length significance defined as the significance of the distance between the primary vertex and displaced vertex, (d) the energy fraction, defined as the energy of the tracks in the displaced vertex relative to the energy of all tracks reconstructed within the jet, (e) the invariant mass and (f) the number of tracks associated to the vertex. From [65].

decay multiplicity and a decay distance similar to b-hadrons'. Such a set was introduced to improve further the c-tagging performance of the algorithm [69]. It includes the invariant mass and the energy of the tracks associated to the secondary vertex, as well as their rapidity with respect to the jet axis and with respect to the heavy-flavour hadron trajectory.
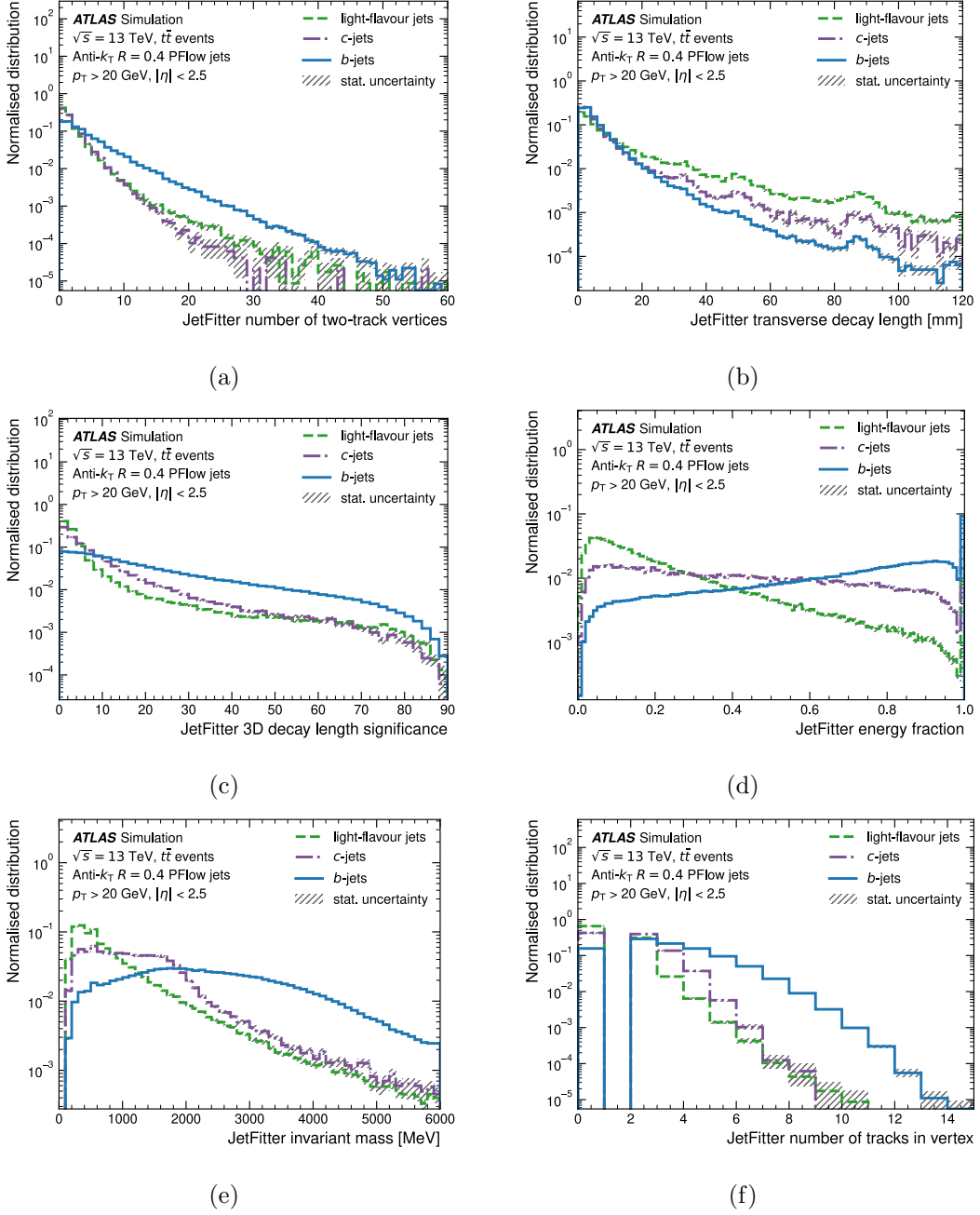
Figure 3.15: Properties of secondary vertices reconstructed by the `JetFitter` algorithm for b-jets, c-jets and light-flavour jets in $t\bar{t}$ simulated events: (a) the number of two-track vertices reconstructed within the jet, (b) the transverse decay length, (c) the average 3D decay length significance, defined as the significance of the average distance between the primary vertex and displaced vertices, (d) the energy fraction, defined as the energy of the tracks in the displaced vertex relative to the energy of all tracks reconstructed within the jet, (e) the invariant mass and (f) the number of tracks associated with the vertex. From [65].

3.2.3   Soft Muon Tagger: `SMT`

As anticipated before, soft muons have an enhanced presence in b-jets, due to the semi-leptonic decays of b-hadrons ($BR(b \to \mu\nu X) \sim 11\%$) and c-hadrons emerging from decays of b-hadrons ($BR(b \to c \to \mu\nu X) \sim 10\%$). `SMT` [69] uses the features of these soft muons to improve b-tagging.

Only combined muons within a cone of $\Delta R < 0.4$ from the jet axis are considered, further subject to selections regarding their p_T, IPs and $|\eta|$. In light-flavour jets, muons meeting these conditions are predominantly those from the decays of prompt W-boson, light hadrons, light and strange mesons, or τ leptons. Misidentified energetic hadrons that reach the MS also constitute a small but sizeable background.

A set of variables has been defined to identify muons related to a heavy-flavour hadron semi-leptonic decay. Nowadays, these variables have been tested to train some yet uncalibrated versions of the high-level taggers. They include the ΔR between the muon and the jet axis, d_0, and $p_T^{\rm rel}$, the projection of the muon's p_T onto the jet axis (Fig. 3.17). The $p_T^{\rm rel}$ is typically larger for particles from the decays of heavy particles, so it's useful for spotting muons from semi-leptonic b-hadrons' decays, but less effective for those from c-hadrons.

3.2.4   The `DL1` high-level algorithms

The distinct information provided by low-level taggers, which target specific features of heavy-flavour jets, is used to train the multivariate high-level algorithms. In Fig. 3.5, two series of high-level taggers are included: `MV2` [75] and `DL1` [65]. The former, whose members consist of BDTs, was used in many ATLAS analyses during Run-2. As `MV2` is outperformed by `DL1` and was not used for this research, it is not further detailed.

The `DL1` algorithms are fully connected, multi-layer feed-forward NNs. They assign three scalar outputs to each jet, p_b, p_c, and p_u, which signify the probabilities that they evaluate for the jet being a b, c, or light-flavour jet. These probabilities are then combined into two distinct discriminant variables, D_b and D_c, as in

$$D_{b/c} = \ln \left(\frac{p_{b/c}}{f_{c/b}p_{c/b} + (1 - f_{c/b})p_u} \right), \tag{3.3}$$

where f_c and f_b are adjustable parameters governing the trade-off between the rejection rates of the two classes of background jets[9]. Selections based on D_b and D_c define regions, allowing for the classification of jets as b or c-tagged with progressively increasing confidence.

The algorithms are trained on a simulated *hybrid* sample, composed 70% of $t\bar{t}$ events and 30% of Z' events. Both p_T and $|\eta|$ of jets are included among the training variables, allowing the NN to learn their correlations with the low-level taggers' outputs. However,

---

[9]*i.e.*, c-jets and light-flavour jets when it comes to b-tagging through D_b, and b-jets and light-flavour jets when it comes to c-tagging through D_c.
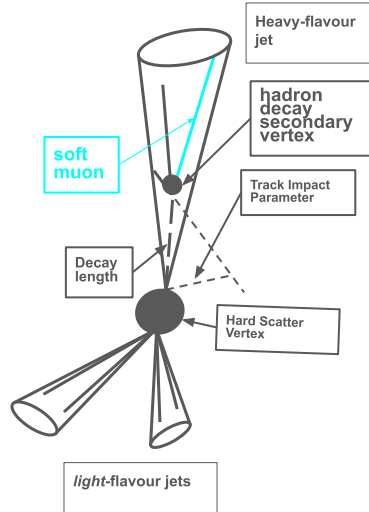
Figure 3.16: *b*-hadrons and *c*-hadrons originating from *b*-hadrons' decays often undergo semi-leptonic decays. The resulting leptons help identify the *b*-jets that contain them.

the training dataset is resampled [10] to ensure that the p_{T} and η distributions of all jet flavours are uniform, preventing the models from identifying their flavours predominantly based on the kinematic distributions of the signal and background jets.

The two algorithms most relevant from the standpoint of this research, `DL1r` and `DL1d`, are described in the following paragraphs.

The `DL1r` algorithm: `DL1r`, where the `r` stands for `RNNIP`, uses as inputs the `IP2D`, `IP3D`, `RNNIP`, `SV1` and `JetFitter` outputs, along with the jet p_{T} and $|\eta|$. The network architecture is schematized in Fig. 3.18. The full list of training variables is found in Tab. 3.1, while Tab. 3.2 enumerates the hyperparameters that shape the NNs architecture and define its training campaign. Fig. 3.19 displays the D_b and D_c distributions evaluated with the `DL1r` algorithm on the $t\bar{t}$ simulated sample.

The `DL1d` algorithm: `DL1d` substitutes the `RNNIP` outputs within `DL1r` with a joint architecture that integrates `DIPS` and `DL1`, as shown in Fig. 3.20. Tracks are processed with a `DIPS` block that estimates p_b, p_c, and p_u. The last hidden layer of the block is also concatenated[11] to a vector created by processing the same jet features used by `DL1` through one layer. The resulting vector is then processed through the NN $\mathcal{U}$, which yields additional estimates for p_b, p_c, and p_u.

---

[10]More information about an analogous resampling procedure is provided in Sec. 4.3.1.

[11]In this context, concatenating two vectors means joining them together into a higher-dimensional single vector.
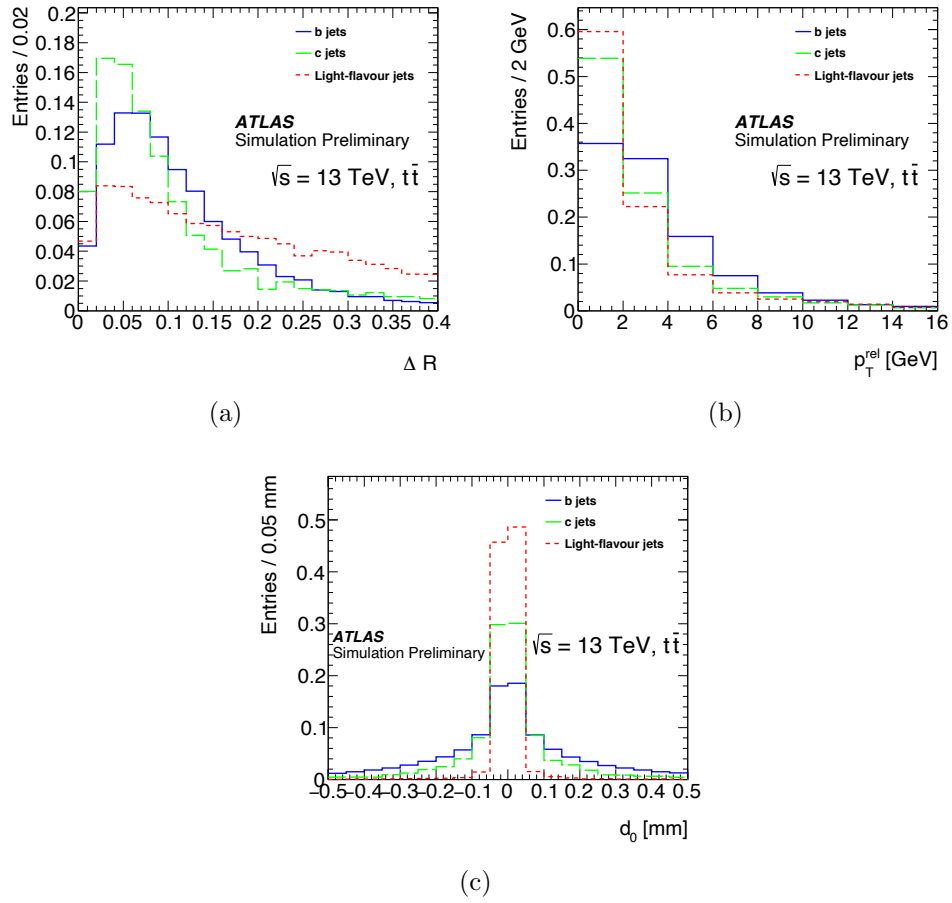
(a)

(b)

(c)

Figure 3.17: ΔR (a), $p_{\mathrm{T}}^{\mathrm{rel}}$ (b) and d_0 (c) normalised distributions for reconstructed muons associated to b-jets (blue), c-jets (green) and light-flavour jets (red). From [69].
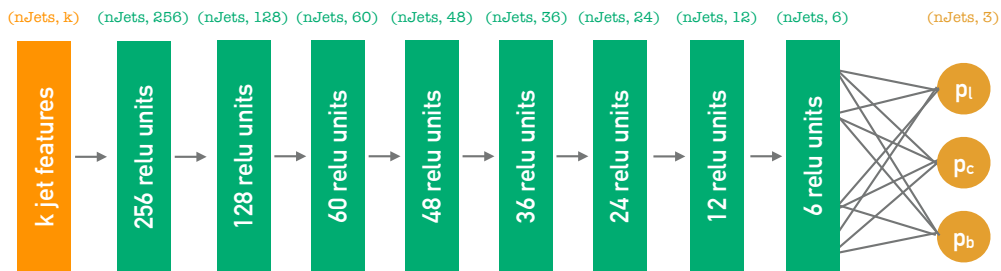

Figure 3.18: Architecture of the `DL1r` algorithm. Image from the ATLAS internal flavour-tagging documentation web site.

Input	Variable	Description
Kinematics	p_T	Jet p_T
	η	Jet $\lvert\eta\rvert$
IP2D, IP3D	$\log(P_b/P_{\text{light}})$	Likelihood ratio of the b-jet to light-flavour jet hypotheses
	$\log(P_b/P_c)$	Likelihood ratio of the b-jet to c-jet hypotheses
	$\log(P_c/P_{\text{light}})$	Likelihood ratio of the c-jet to light-flavour jet hypotheses
RNNIP	P_b	b-jet probability
	P_c	c-jet probability
	P_{light}	light-flavour jet probability
SV1	$m(\text{SV})$	Invariant mass of tracks at the secondary vertex assuming pion mass
	$f_E(\text{SV})$	Jet energy fraction of the tracks associated with the secondary vertex
	$N_{\text{TrkAtVtx}}(\text{SV})$	Number of tracks used in the secondary vertex
	$N_{\text{2TrkVtx}}(\text{SV})$	Number of two-track vertex candidates
	$L_{xy}(\text{SV})$	Transverse distance between the primary and secondary vertices
	$L_{xyz}(\text{SV})$	Distance between the primary and secondary vertices
	$S_{xyz}(\text{SV})$	Distance between the primary and secondary vertices divided by its uncertainty
	$\Delta R(\vec{p}_{\text{jet}}, \vec{p}_{\text{vtx}})(\text{SV})$	ΔR between the jet axis and the direction of the secondary vertex relative to the primary vertex.
JetFitter	$m(\text{JF})$	Invariant mass of tracks from displaced vertices
	$f_E(\text{JF})$	Jet energy fraction of the tracks associated with the displaced vertices
	$\Delta R(\vec{p}_{\text{jet}}, \vec{p}_{\text{vtx}})(\text{JF})$	ΔR between the jet axis and the vector sum of momenta of all tracks attached to displaced vertices
	$S_{xyz}(\text{JF})$	Significance of the average distance between PV and displaced vertices
	$N_{\text{TrkAtVtx}}(\text{JF})$	Number of tracks from multi-prong displaced vertices
	$N_{\text{2TrkVtx}}(\text{JF})$	Number of two-track vertex candidates (prior to decay chain fit)
	$N_{\text{1-trk vertices}}(\text{JF})$	Number of single-prong displaced vertices
	$N_{\geq\text{2-trk vertices}}(\text{JF})$	Number of multi-prong displaced vertices
	$L_{xyz}(2^{\text{nd}})(\text{JF})$	Distance of 2nd vertex from PV
	$L_{xy}(2^{\text{nd}})(\text{JF})$	Transverse displacement of the 2nd vertex
	$m_{\text{Trk}}(2^{\text{nd}})(\text{JF})$	Invariant mass of tracks associated with the 2nd vertex
	$E(2^{\text{nd}})(\text{JF})$	Energy of the tracks associated with the 2nd vertex
	$f_E(2^{\text{nd}})(\text{JF})$	Jet energy fraction of the tracks associated with the 2nd vertex
	$N_{\text{TrkAtVtx}}(2^{\text{nd}})(\text{JF})$	Number of tracks associated with the 2nd vertex
	$\eta_{\text{trk}}^{\text{min,max,avg}}(2^{\text{nd}})(\text{JF})$	Min., max. and avg. pseudorapidity of tracks at the 2nd vertex

Table 3.1: Input variables used by the `DL1r` algorithm.

Hyperparameter	Value
Number of input variables	31
Number of hidden layers	8
Number of nodes [per layer]	[256, 128, 60, 48, 36, 24, 12, 6]
Learning rate	0.01
Training batch size	15 000
Activation function	ReLu
Number of training epochs	200
Free (trainable) parameters	59 275
Fixed parameters	1 140
Training sample size	22 M jets

Table 3.2: List of optimised hyperparameters shaping the `DL1r` architecture and defining its training campaign.
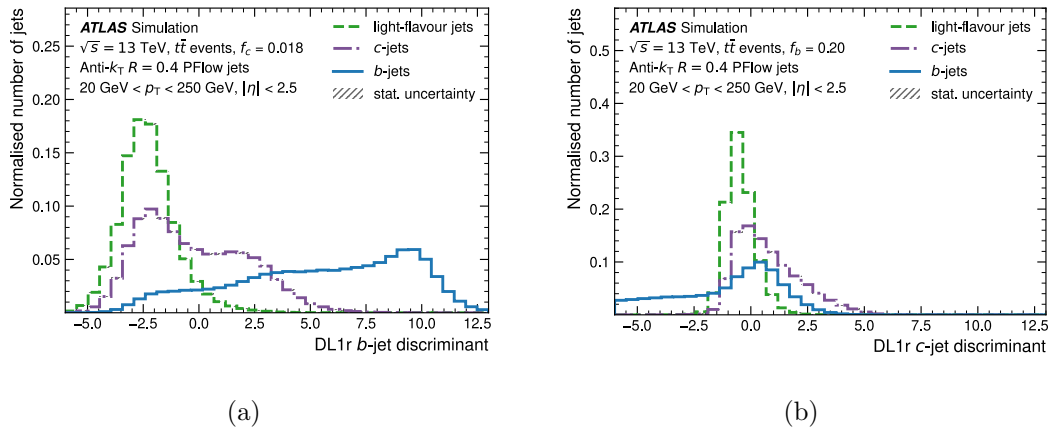

Figure 3.19: D_b and D_c distributions evaluated with the `DL1r` algorithm on the $t\bar{t}$ simulated sample. From [65].
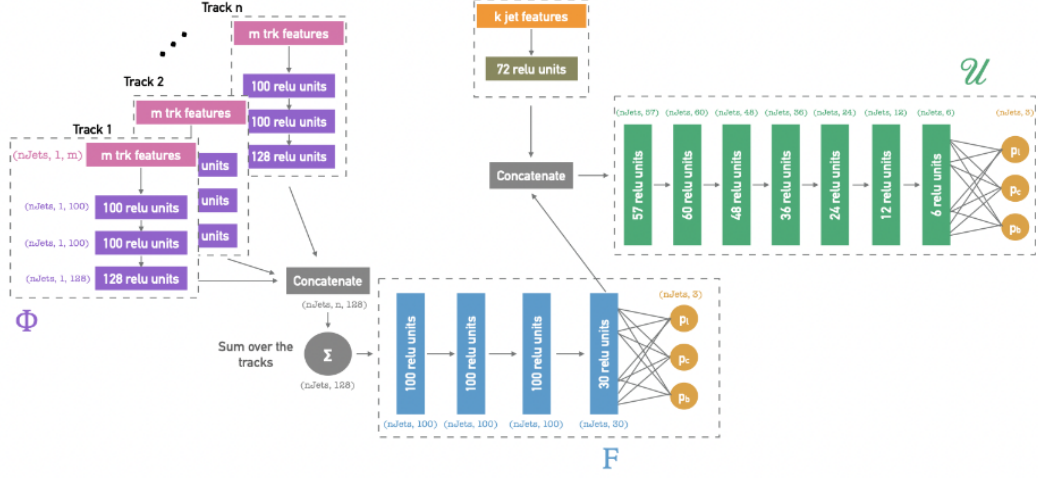
Figure 3.20: Architecture of the `DL1d` algorithm. Image from the ATLAS internal flavour-tagging documentation web site.

Both the p_b, p_c, and p_u estimates have a dedicated loss. That assigned to the `DIPS` block is only aware of the features of the tracks, while the loss of the $\mathcal{U}$ network is affected by both the track and jet features. These two terms are then merged into a single loss function, which is employed to perform full back-propagation, simultaneously optimizing the $\mathcal{U}$ and `DIPS` parameters. The `DIPS` loss allows to evaluate the standalone `DIPS` performance and helps optimise the related part of the network. The outputs of the $\mathcal{U}$ network are then combined in the D_b and D_c discriminant variables. Their distributions, evaluated on the $t\bar{t}$ simulated sample, are shown in Fig. 3.21.

3.2.5 Performance

The performance of jet flavour-tagging algorithms can be graphically represented through Receiver Operating Characteristics (ROC) curves. These curves plot the efficiency of correctly tagging signal jets against the rejection rates of background jets. These values are derived by selecting thresholds in the D_b and D_c distributions and then calculating the proportion of signal and background jets above it. These proportions are the flavour-tagging efficiencies ϵ_f of the algorithm at the threshold:

$$\epsilon_f = \frac{\text{number of jets of flavour } f \text{ above the threshold}}{\text{total number of jets of flavour } f}. \tag{3.4}$$

Each point on the curve corresponds to a specific threshold and, thus, to particular jet flavour-tagging efficiencies. The signal jet efficiency of a given threshold is also called the
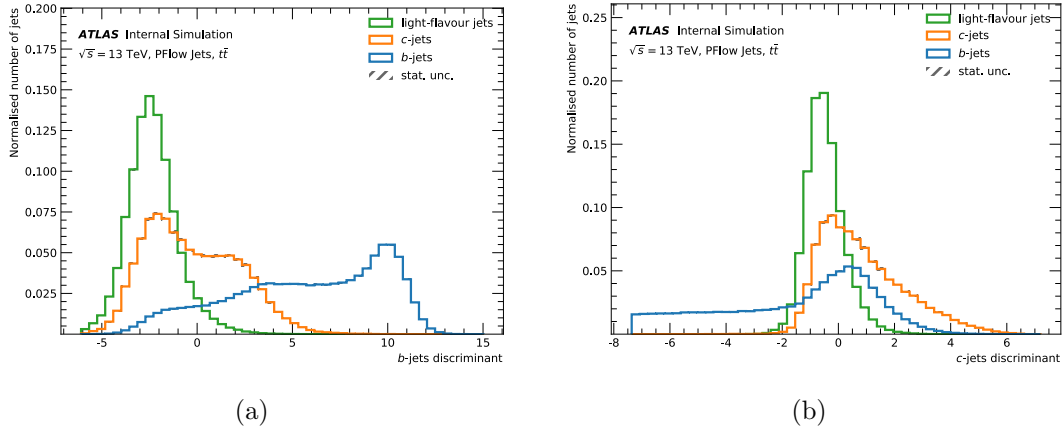
Figure 3.21: D_b and D_c distributions evaluated with the `DL1d` algorithm on $t\bar{t}$ simulated events.

Working Point (WP). The background rejection rates, displayed on the y axis, are computed as the inverse of the background efficiencies. A background rejection rate of r indicates that, in a random set of r background jets, on average, one is *mistagged*, meaning it is erroneously identified as signal.

Fig. 3.22 shows b-tagging ROC curves for the `RNNIP`, `DL1r` and `DL1d` algorithms evaluated on jets from the $t\bar{t}$ simulated sample. Fig. 3.22a compares the performance of `DL1r` and `RNNIP` evaluated on simulated Run-2 events. Two algorithms named `SVKine` and `JFKine` also appear in the plot. These are sub-networks of `DL1` that I developed to assess the standalone jet flavour-tagging performance of the `SV1` and `JetFitter` algorithms. They are described in full detail in the next section. The ROC curves clearly illustrate the benefit of combining the low-level information through high-level algorithms: `DL1r` achieves background rejections up to 4 times larger than the standalone low-level taggers. Fig. 3.22b compares the performance of `DL1r` and `DL1d` evaluated on simulated Run-3 events. `DL1d` surpasses the background rejection rates achieved with `DL1r` of factors up to 30%.

3.3   Developing `SV1` with `SVKine` (and `JFKine`)

A consequence of the two-stage approach characterizing the `DL1` algorithms is that all the low-level tagger algorithms must be maintained when a change occurs in one of the subsequent steps involved in the reconstruction of the detector-level and physics objects that they use.

When I joined the ATLAS Collaboration, a major upgrade of Athena was ongoing, involving substantial changes in the object reconstruction procedure. The effect of these changes on the jet flavour-tagging algorithms had to be assessed. When deemed necessary,
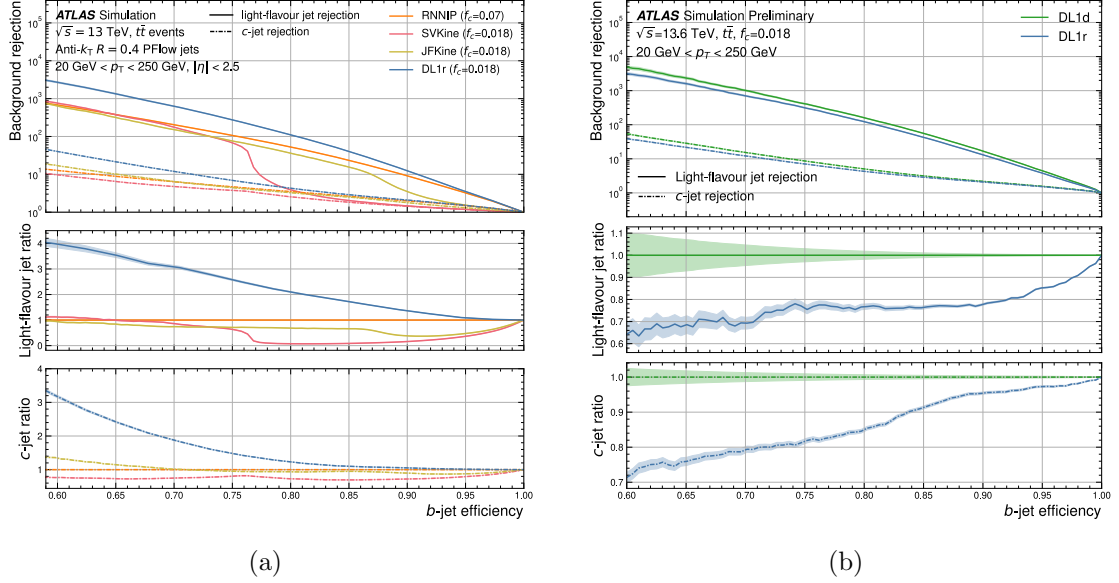
<div style="text-align:center">(a)   (b)</div>

Figure 3.22: (a) *b*-tagging ROC curves of the `DL1r`, `RNNIP`, `SVKine` and `JFKine` algorithms evaluated on simulated jets from the $t\bar{t}$ sample. The statistical uncertainties of the rejection rates are calculated using binomial uncertainties and are indicated as coloured bands. The lower two panels show the ratio of the light-flavour jet rejection and the *c*-jet rejection of the algorithms to `RNNIP`. `SV1` and `JetFitter` have secondary-vertex-finding efficiencies of approximately 80% and 90%, respectively; this causes the rapid growth of their light-flavour jet rejection rates around these values. Figure produced by me and published in [65]. (b) The light-flavour jet (solid line) and *c*-jet (dashed line) rejection rates for the `DL1r` and `DL1d` algorithms, evaluated on $t\bar{t}$ events simulated with the LHC Run-3 conditions. From [64].

a re-optimization of the parameters defining the flavour-tagging algorithms was performed.

`SVKine` and `JFKine` were developed in this context. They are NNs trained only on the variables listed under the sections Kinematics, `SV1` and `JetFitter` of Tab. 3.1. Their scope was to guarantee a fast turnaround in the optimization process of `SV1` and `JetFitter` for the new ATHENA release. The hyperparameters shaping the architecture of the NNs and defining their training campaign are enumerated in Tab. 3.3.

Hyperparameter	Value
Number of input variables	10/17 `SVKine`/`JFKine`
Number of hidden layers	4
Number of nodes [per layer]	[36, 24, 12, 6]
Learning rate	0.01
Training batch size	15 000
Activation function	ReLu
Number of training epochs	200
Training sample size	9.5 M jets

Table 3.3: List of hyperparameters that shape the architecture of the `SVKine` and `JFKine` algorithms and define their training campaign.

`SVKine` and `DL1r` were used in concert to test the benefits of relaxing the cut on `BVrtScore`, applied in the workflow of the `SV1` algorithm to discard two-track vertices not compatible with the decay of an heavy-flavour hadron (Sec. 3.2.3). The value of such cut, historically, was tuned before developing the `DL1r` algorithm. Relaxing it produces two concurring effects: an increased proportion of reconstructed "good" secondary vertices, composed of tracks emerging from the decay of heavy-flavour hadrons, at the cost of increasing also the proportion of "bad" secondary vertices, composed of tracks produced in other processes. The optimization aimed at identifying the value of the cut that would maximise the jet flavour-tagging performance of `DL1r`.

Two WPs of the Single Secondary Vertex Finding (SSVF) step of the `SV1` algorithm, referred to as *Standard* and a *Loose*, were defined. They correspond to cut values on `BVrtScore` (Eq. 3.2) of 0.015 and 0.005, respectively. Fig. 3.23 illustrates the features of the secondary vertices reconstructed with the *Loose* and *Standard* `SV1` WP, as well as those reconstructed by `JetFitter`. Three figures of merit were proposed to characterize the quality of the vertexing performance, all relying on the true information about the process producing the simulated reconstructed tracks:

- the *recall*: the ratio of tracks associated to the secondary vertex and produced by decay products of heavy-flavour hadrons, over the total number of tracks produced by decay products of heavy-flavour hadrons;
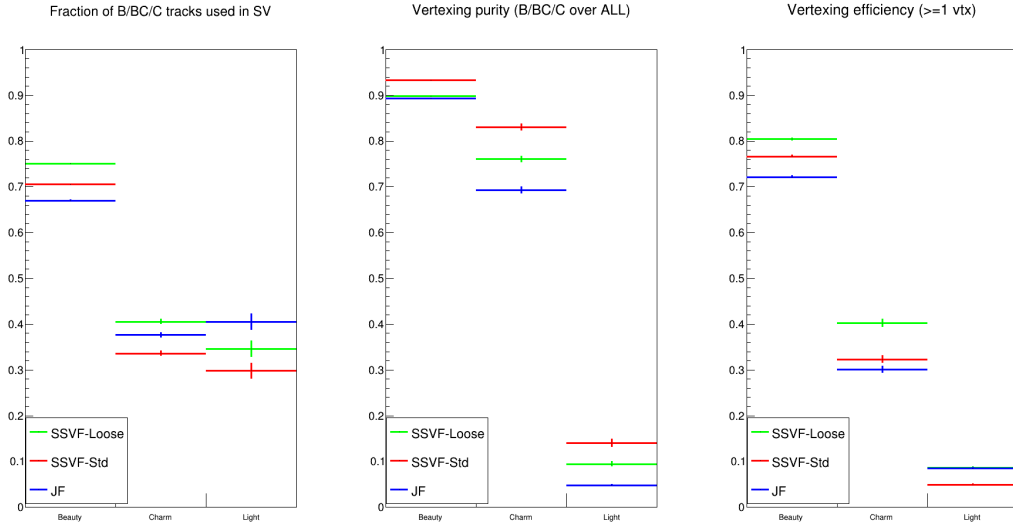
Figure 3.23: Recall, vertex purity and vertexing efficiency of secondary vertices reconstructed with the *Loose* and *Standard* working points of the `SV1` algorithm, and with `JetFitter`. In this figure, B, BC and C tracks indicate tracks known at the truth level to originate from the decay of *b*-hadrons, *c*-hadrons produced in *b*-hadrons decays, and other types of *c*-hadrons respectively. JF stands for `JetFitter`. Loosening the cut on `BVrtScore` increases recall and vertexing efficiency for all the jet flavours and decreases vertex purity.

- the *vertex purity*: the ratio of tracks associated to the secondary vertex and produced by decay products of heavy-flavour hadrons, over the total number of tracks associated to the secondary vertex;

- the *vertexing efficiency*: the fraction of jets with at least a reconstructed secondary vertex.

Choosing the *Loose* WP increases recall and vertexing efficiency for all the jet flavours, but reduces vertex purity.

The increase in vertexing efficiency achieved when loosening the cut is balanced by the lower quality of the reconstructed vertices. "Good" secondary vertices were defined as those formed by at least 50% of tracks from the decays of heavy-flavour hadrons. Figure 3.24 shows the "good" vertexing efficiency as a function of the `BVrtScore` cut value for the different jet flavours. The efficiency measured in light-flavour jets decreases more for tighter cuts when compared to heavy-flavour jets.

`SVKine` was trained multiple times utilizing the `SV1` variables obtained when applying different cut values on `BVrtScore`. This was done to assess how the interplay between varying
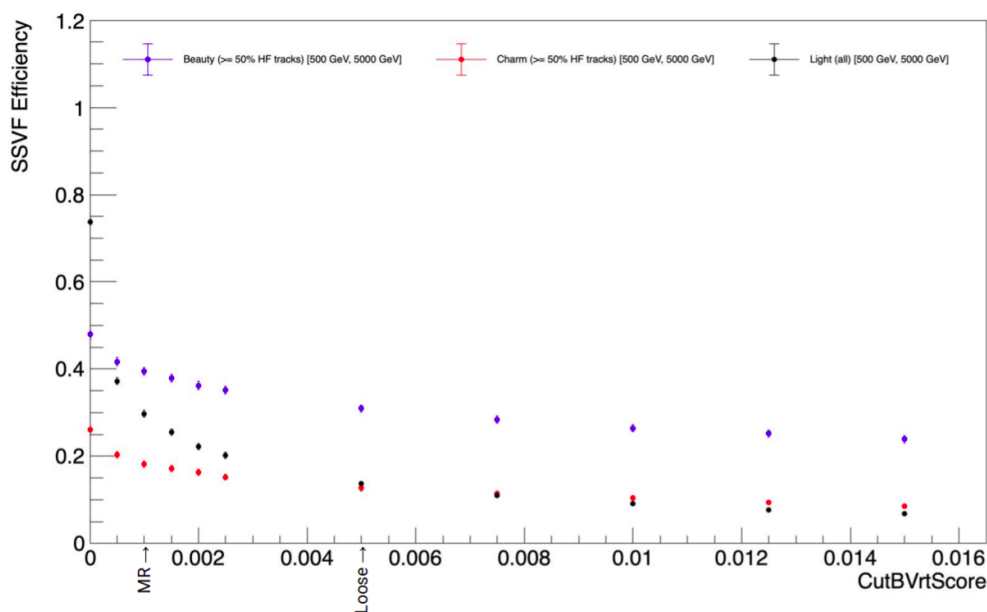
Figure 3.24: "Good" vertexing efficiency, *i.e.*, vertexing efficiency for vertices composed of at least 50% of tracks stemming from the decay of a heavy-flavour hadron, as a function of the `BVrtScore` cut value in b (blue), c (red) and light-flavour jets (black). The Standard and Loose WPs correspond to cut values of 0.015 and 0.005, respectively.
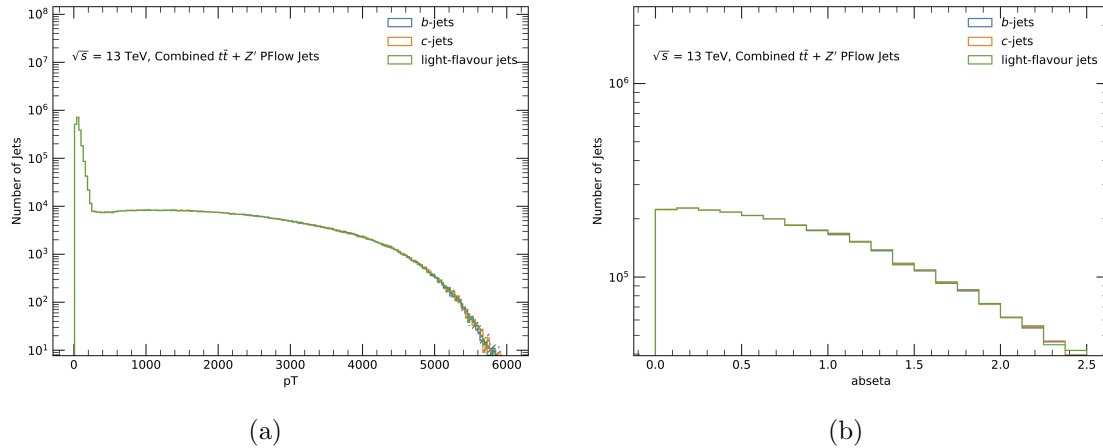
(a)   (b)

Figure 3.25: p_{T} and $|\eta|$ distributions of jets in the hybrid $t\bar{t}$ and Z' sample used to train `SVKine` and `JFKine`. The c-jets and light-flavour jets spectra are resampled to match the b-jets'.

vertexing efficiency, purity and recall affects the jet flavour-tagging information contained in the reconstructed vertices, when harnessed by a NN capable of exploiting the correlations among the various vertex features. The training is performed on a hybrid $t\bar{t}$ and Z' sample, resampled to match the p_{T} and $|\eta|$ distributions of c-jets and light-flavour jets to the b-jets' (Fig. 3.25).

Fig. 3.26 shows the b-tagging ROC curves computed with the different versions of `SVKine`, when it is evaluated on the simulated $t\bar{t}$ and Z' samples. The mark of the concurring effects obtained when varying the `BVrtScore` cut is clearly visible in the behaviour of the light-flavour jet rejection rate as a function of the b-tagging efficiency. ROC curves evaluated using tighter cut values achieve better light-flavour jet rejection rates at lower efficiency values. However, these rates drop sooner when approaching higher efficiencies compared to algorithms using looser cut values. These drops in rejection rates at increasing efficiencies for loosening `BVrtScore` cut values, particularly evident when looking at the curve evaluated on the $t\bar{t}$ sample, indicate that when `SV1` is not capable of providing secondary vertices, no discriminant information is available. In other words, when `SVKine` is evaluated at a b-tagging efficiency greater than the proportion of b-jets in which a secondary vertex is reconstructed, it identifies as b-jets also those with no reconstructed secondary vertex, which cannot be distinguished from the light-flavour jets by the algorithm. This is clear when comparing `SVKine`'s D_b distributions evaluated on the $t\bar{t}$ sample using the standard cut of `BVrtScore` = 0.015 (Fig. 3.27a) and using no cut at all (Fig. 3.27b). The former figure displays distributions in which two distinct zones can be identified. Jets with reconstructed secondary vertices are found at D_b values larger than 0. This region is populated mostly by b-jets, followed by c-jets and a very small amount of light-flavour jets. Jets with no reconstructed secondary vertex populate the region with $D_b < 0$ instead, which is dom-

inated by light-flavour jets. b-tagging WPs of `SVKine` with ϵ_b values greater than those corresponding to the drops in the light-flavour jet background rejection rates use thresholds set in the $D_b < 0$ zone. Any of these thresholds gives a definition of b-tagged jet which applies to significant fractions of the background jets as well, greatly deteriorating the rejection rate. In Fig. 3.27b, no `BVrtScore` cut is imposed on the reconstructed secondary vertex. The result is that most of the background jets contain at least a secondary vertex as well (Fig. 3.24). In this case, the D_b distributions are more similar for the different jet flavours and the separation between the zones of the distributions populated by jets with and without reconstructed vertices is less evident.