Lund Plane to Bloch (LP2B) Encoding for Object and Polarization Tagging with Quantum Jet Substructure
ML4Jets - Vienna 18/09/2026

Fabrizio Napolitano, Tommaso Tedeschi, Livio Fanò, Luca Della Penna

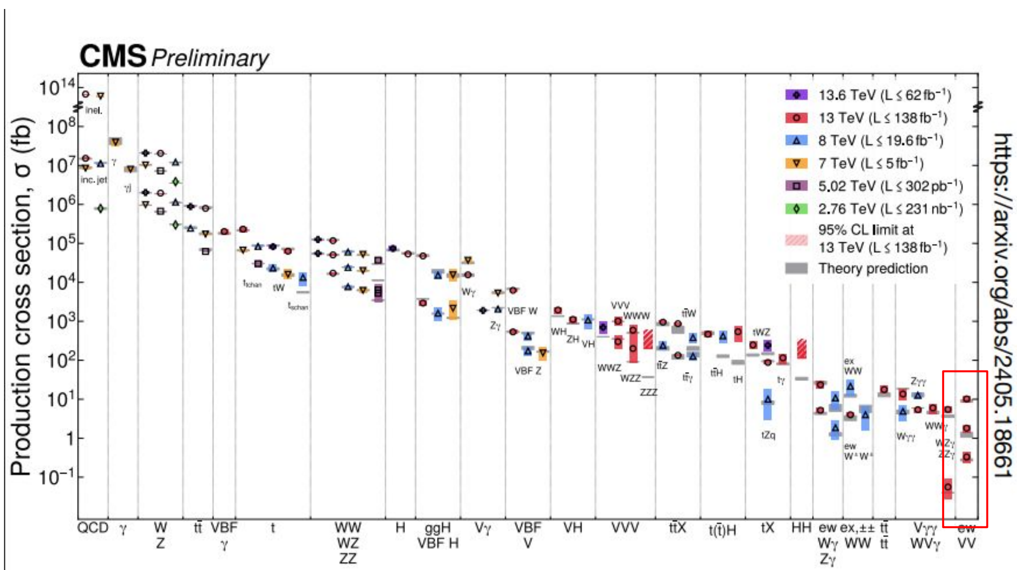
https://link.springer.com/article/10.1140/epjc/s10052-026-16142-9

Some general considerations as guidelines:

- building a tagger for object/polarization from jet substructure needs to limits the dependence from the production process, i.e, we want to learn from X $\rightarrow$ jet and not from pp $\rightarrow$ XY $\rightarrow$ jet Y
- systematics can be large under different theoretical assumptions (Pythia vs Herwig vs Sherpa vs Generator X) and different inputs [1]
- AUC is not the only thing to look, hopefully a tagger should be resilient. [2]
-

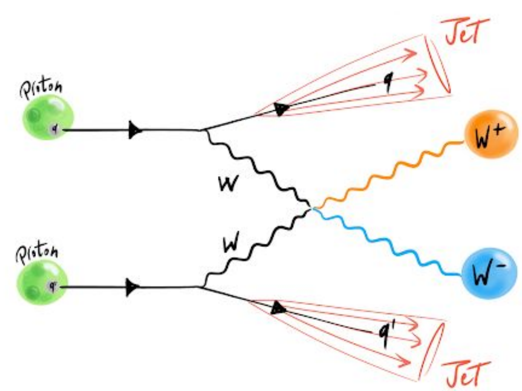
Some particular consideration for polarized LHC analyses (why we included the polarization task?):

- Polarization tagging is now an active field for VBS: test EWSB and possible BSM effects [3] [4]
- No available measurements of polarized VBS-like processes in semileptonic/fully hadronic channel from CMS/ATLAS collaborations



VBS

rarest observed EWK process at LHC! (~fb)



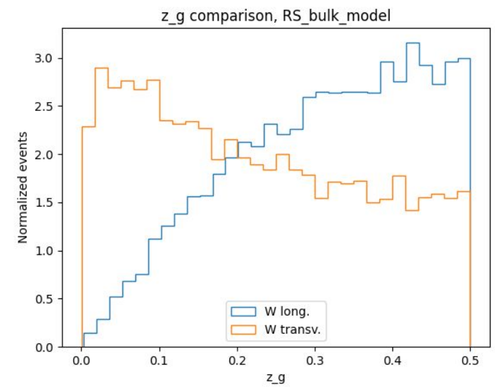
Polarization Tagging: (old) Lund Plane?

- It decluster a jet by undoing the last step of the C/A clustering algorithm to have two subjets
- at each step, define the following variables and follow the harder subjet

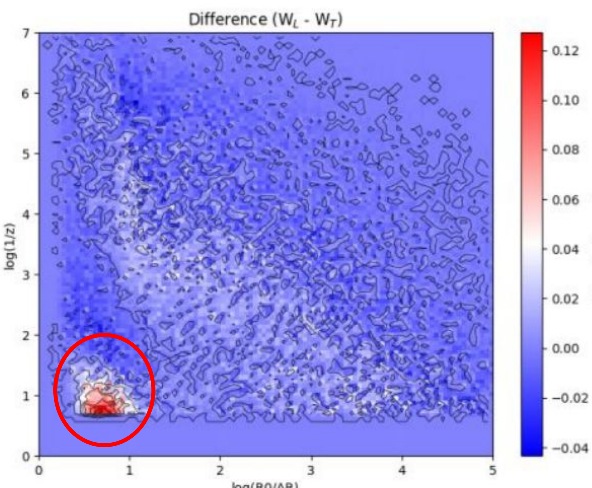
$$\Delta \equiv \Delta R_{ij}, \qquad k_t \equiv p_{tj}\Delta, \qquad m^2 \equiv (p_i + p_j)^2$$

$$z \equiv \frac{p_{tj}}{p_{ti} + p_{tj}}, \qquad \kappa \equiv z\Delta, \qquad \psi \equiv \tan^{-1}\frac{y_j - y_i}{\phi_j - \phi_i}$$

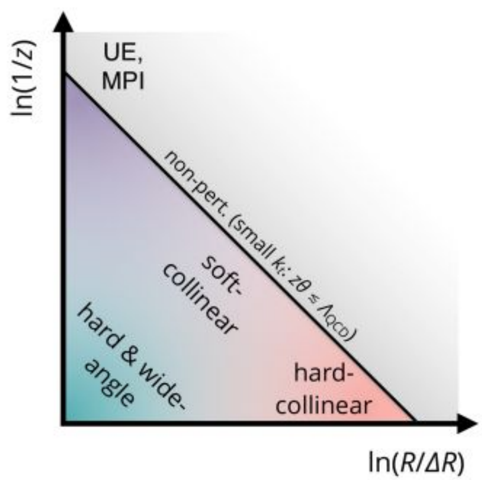
- back to step 1 with the harder branch of step 2 until we reach the end of the tree



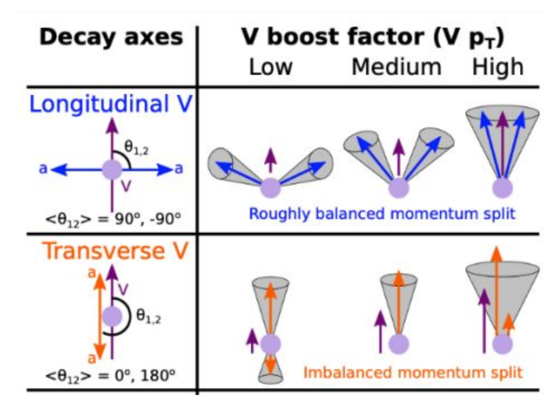
Lund Plane implementation



VBS WWLL excess



POLARIZATION



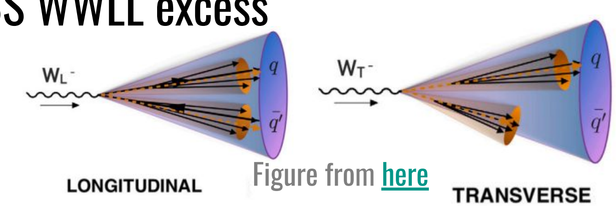


Figure from here

Polarization Tagging: (old) Lund Plane?

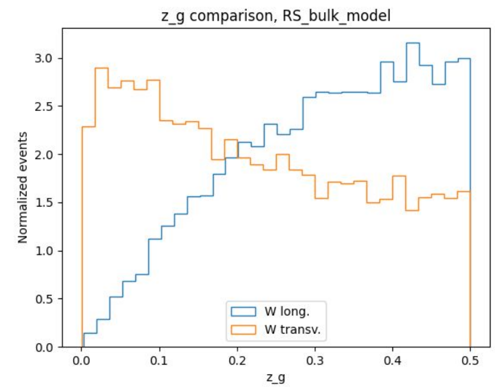
- It decluster a jet by undoing the last step of the C/A clustering algorithm to have two subjets
- at each step, define the following variables and follow the harder subjet

$$\Delta \equiv \Delta R_{ij}, \qquad k_t \equiv p_{tj}\Delta, \qquad m^2 \equiv (p_i + p_j)^2$$

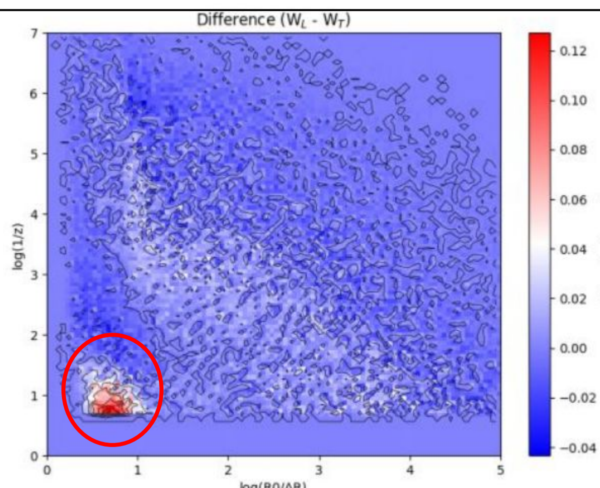
$$z \equiv \frac{p_{tj}}{p_{ti} + p_{tj}}, \qquad \kappa \equiv z\Delta, \qquad \psi \equiv \tan^{-1}\frac{y_j - y_i}{\phi_j - \phi_i}$$

this is the polarization proxy

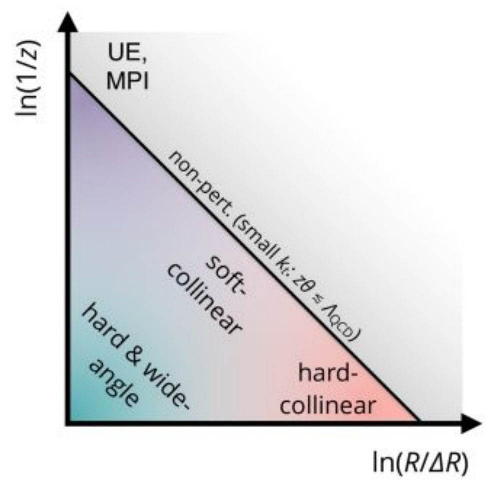
- back to step 1 with the harder branch of step 2 until we reach the end of the tree



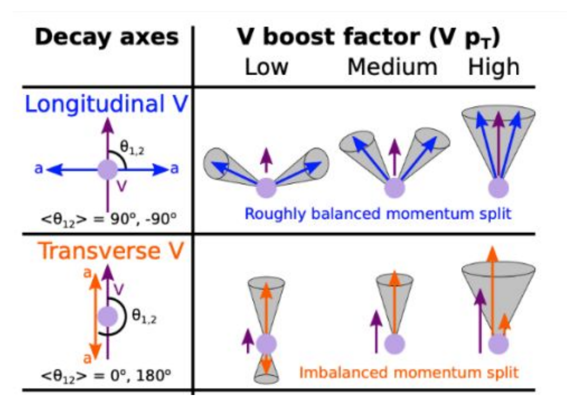
Lund Plane implementation



VBS WWLL excess



POLARIZATION



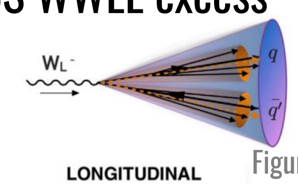
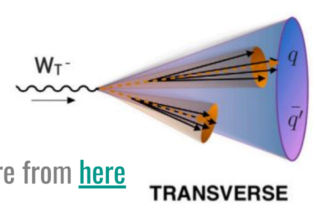


Figure from here



Lund Plane for Polarization Tagging…
and standard object tagging

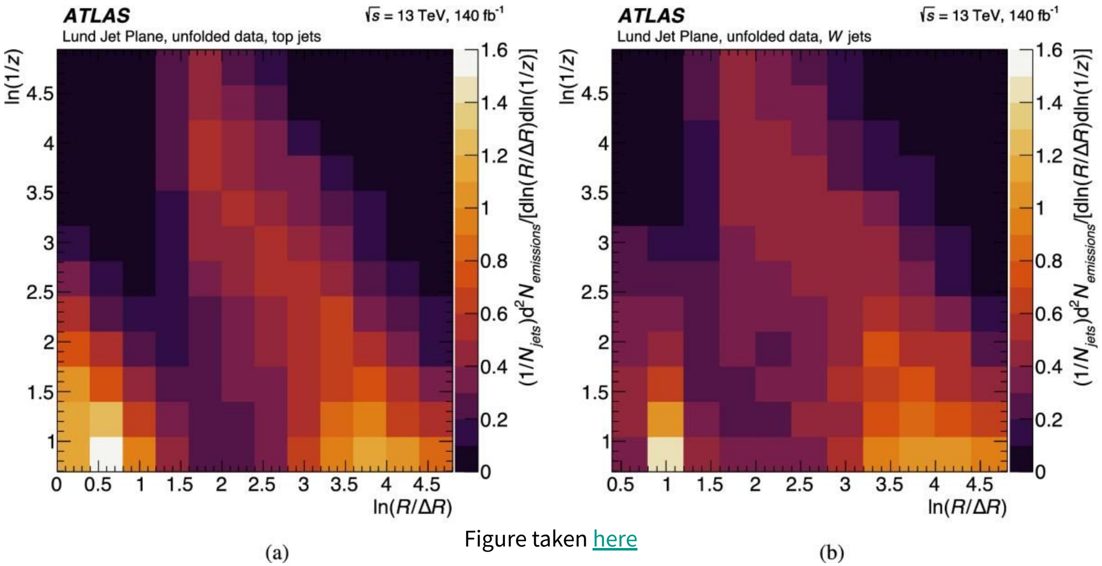
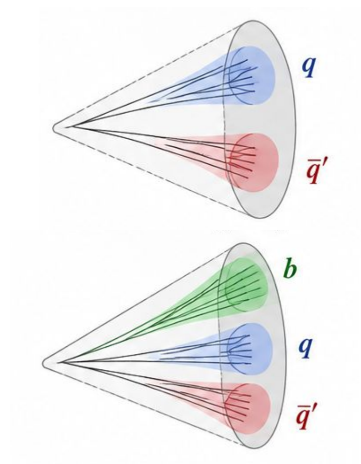


Figure taken here

top/W



Lund Plane is a well known tool already used for a lot of different application in jet substructure, starting from traditional log-likelihood fit to GNN LundNet architecture

Some advantages:

- declustering is purely geometric (C/A)
- For polarization..the umbalance is already encoded in the declustering history (direct link to physic proxy)
- It is Infrared and Collinear Safe (nice to be fulfilled for whatever jet observables)
- It does not contain "raw" particle informations and most of the informations is in the first splits of the declustering (nice to have if we can express jet properties with a lower number of parameters)

can we move to a quantum implementation that can be feasible for near-term NISQ, i.e, with a low number of qubits?

Quick Quantum Computing

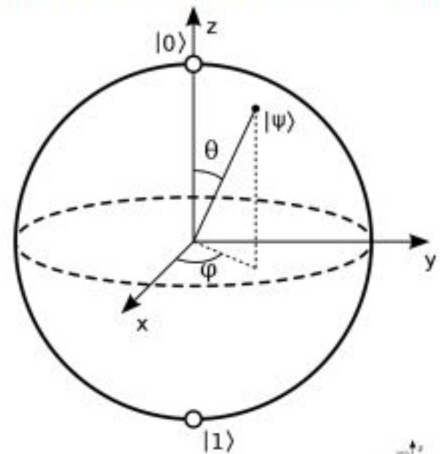
Qubit

$$\mathcal{H} = \mathbb{C}^2$$

Computational basis

$$\{|0\rangle = \begin{pmatrix}1\\0\end{pmatrix}, |1\rangle = \begin{pmatrix}0\\1\end{pmatrix}\}$$

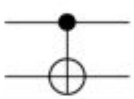
Bloch sphere representation



Arbitrary states can be written as superposition

$$|\psi\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + e^{i\phi}\sin\left(\frac{\theta}{2}\right)|1\rangle$$

Entanglement of qubits

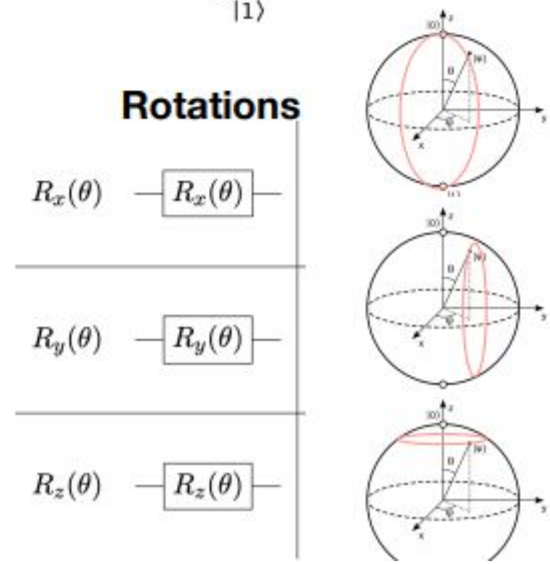


$$\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \otimes |0\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |10\rangle)$$

$$|00\rangle \rightarrow |00\rangle$$
$$|01\rangle \rightarrow |01\rangle$$
$$|10\rangle \rightarrow |11\rangle$$
$$|11\rangle \rightarrow |10\rangle$$

$$\xrightarrow{\text{CNOT}} \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$$

Rotations



LundPlane2Bloch

Lund-Plane has some advantages:

- declustering is purely geometric (C/A)
- For polarization..the umbalance is already encoded in the declustering history (direct link to physic proxy)
- It is Infrared and Collinear Safe (nice to have for whatever jet observable)
- It does not contain raw particle informations and most of the informations is encoded in the first splits of the declustering (nice to have, lower number of parameters)

can we move to a quantum implementation that can be feasible for NISQ…moving from a plane to a sphere?

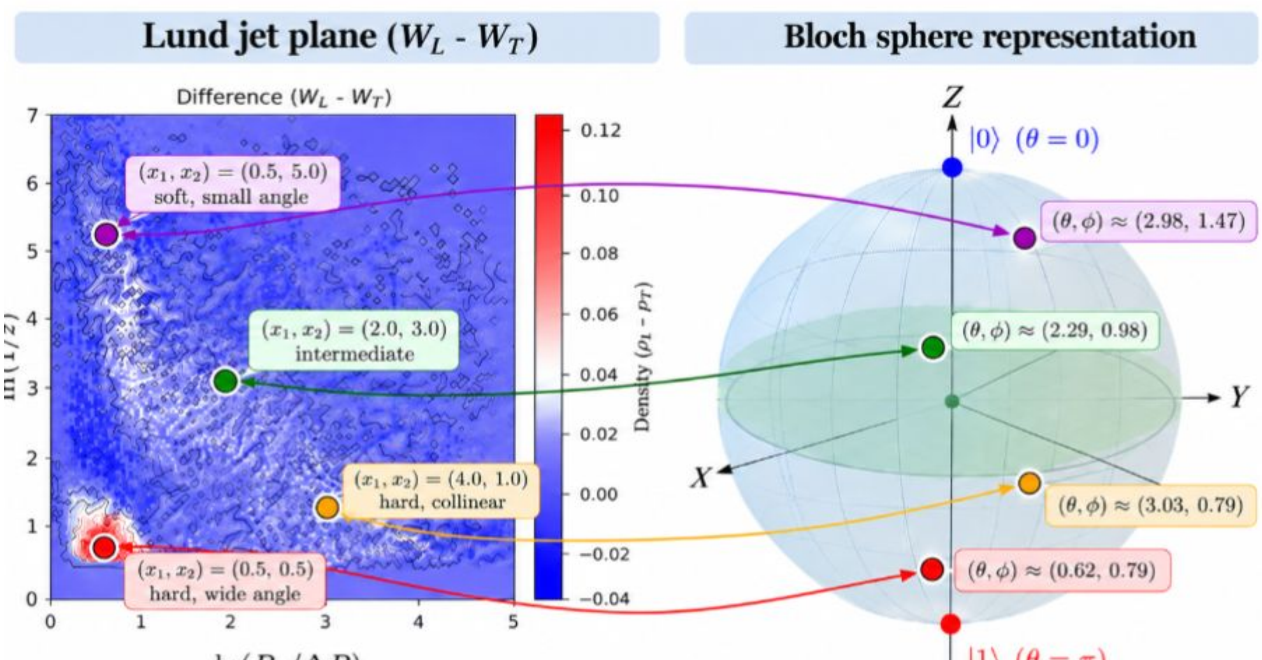


ILLUSTRATION PURPOSE ONLY (FIXED λ, ω=1)

$$r = \sqrt{x_1^2 + x_2^2 + \epsilon}$$

$$\theta = 2\arctan(\lambda_i \cdot r)$$

$$\phi = \omega_i \cdot \mathrm{atan2}(x_2, x_1)$$

$$x_1 = \log\left(\frac{1}{z}\right)$$

$$x_2 = \log(1/R)$$

LundPlane2Bloch

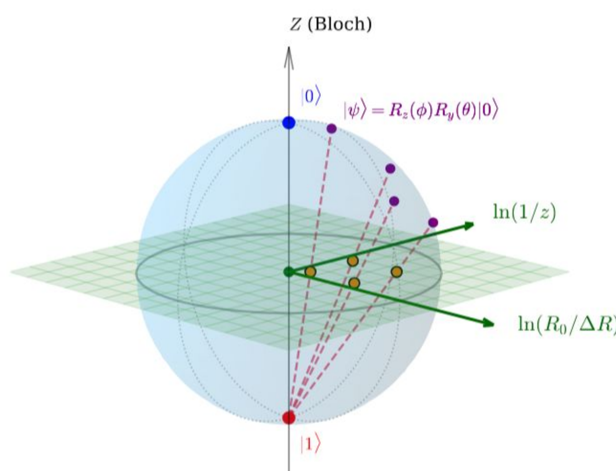
Lund-Plane has some advantages:
- declustering is purely geometric (C/A)
- For polarization..the umbalance is already encoded in the declustering history (direct link to physic proxy)
- It is Infrared and Collinear Safe (nice to have for whatever jet observable)
- It does not contain raw particle informations and most of the informations is encoded in the first splits of the declustering (nice to have, lower number of parameters)

can we move to a quantum implementation that can be feasible for NISQ…moving from a plane to a sphere and without losing these properties?

- Empty nodes are automatically encoded to the North-Pole, such that noise is not introduced and we are still IRC safe and we can handle with sparse data (depth of the jet-tree not a-priori known).
- Trainable parameters in the encoding and specialized for each node of the tree such that we can span the Bloch Sphere to represent QCD phase space

$$r = \sqrt{x_1^2 + x_2^2 + \epsilon}$$
$$\theta = 2\arctan(\lambda_i \cdot r)$$
$$\phi = \omega_i \cdot \mathrm{atan2}(x_2, x_1$$

$$x_1 = \log\left(\frac{1}{z}\right)$$
$$x_2 = \log(1/\Delta R)$$

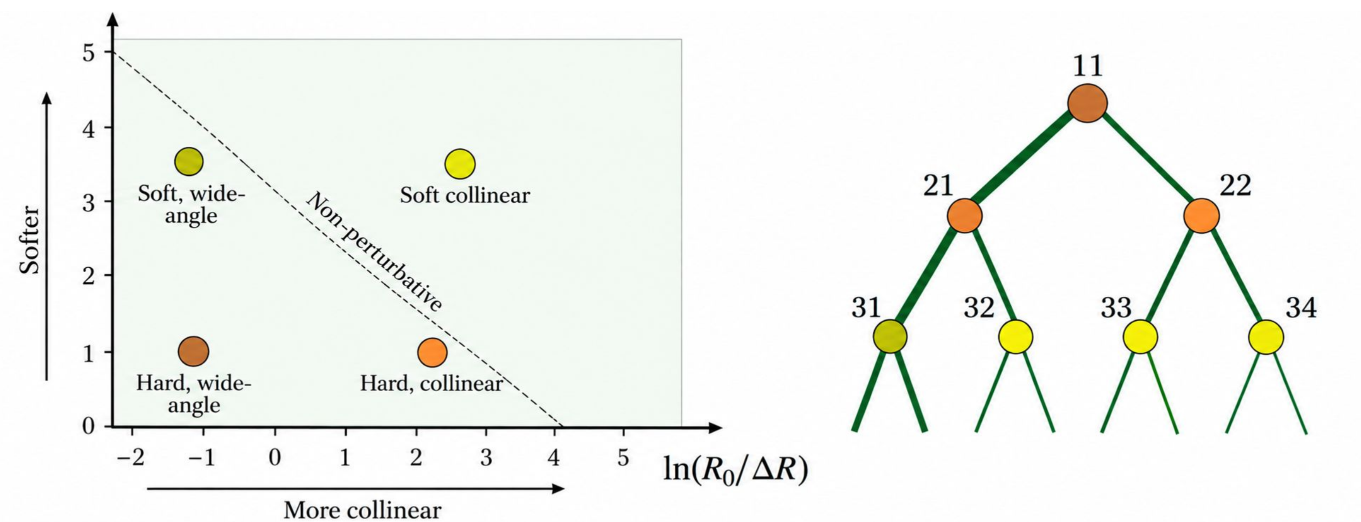


LundPlane2Bloch

Lund-Plane has some advantages:
- declustering is purely geometric (C/A)
- For polarization..the umbalance is already encoded in the declustering history (direct link to physic proxy)
- It is Infrared and Collinear Safe (nice to have for whatever jet observable)
- It does not contain raw particle informations and most of the <u>informations is encoded in the first splits</u> of the declustering (nice to have, lower number of parameters)

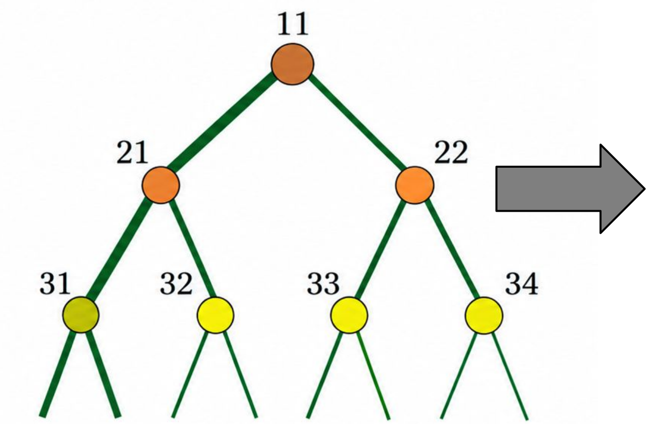
can we move to a quantum implementation that can be feasible for NISQ…moving from a plane to a sphere and without losing these properties?

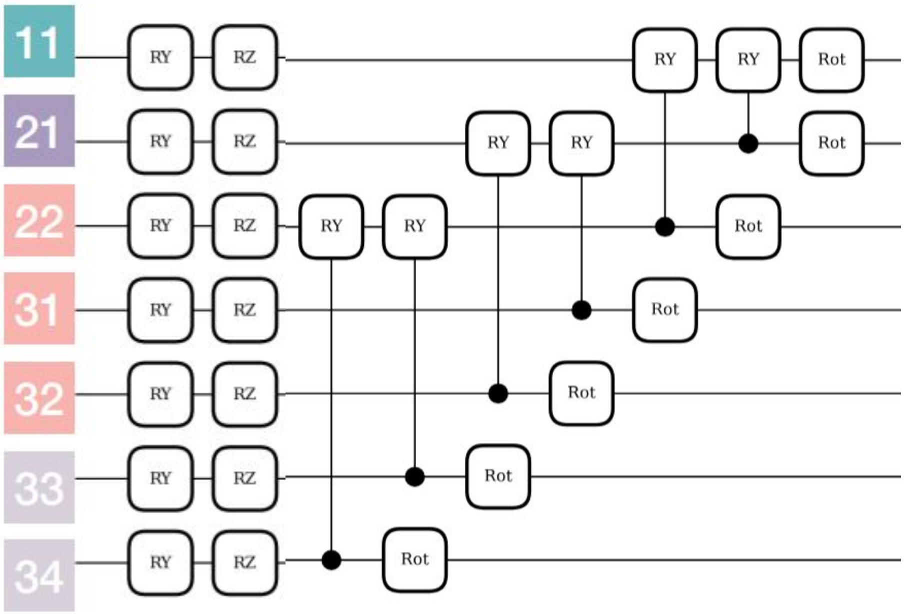
- Truncation of the Lund Tree to D=3 such that information is contained in N=7 qubits. Additionally, lnKt(1) > 1 cut to move upwards the tree the relevant splittings.



Quantum-Tree-Tensor-Network

Building the Quantum ansatz with a tree-tensor-network by reproducing the Lund Tree Graph structure





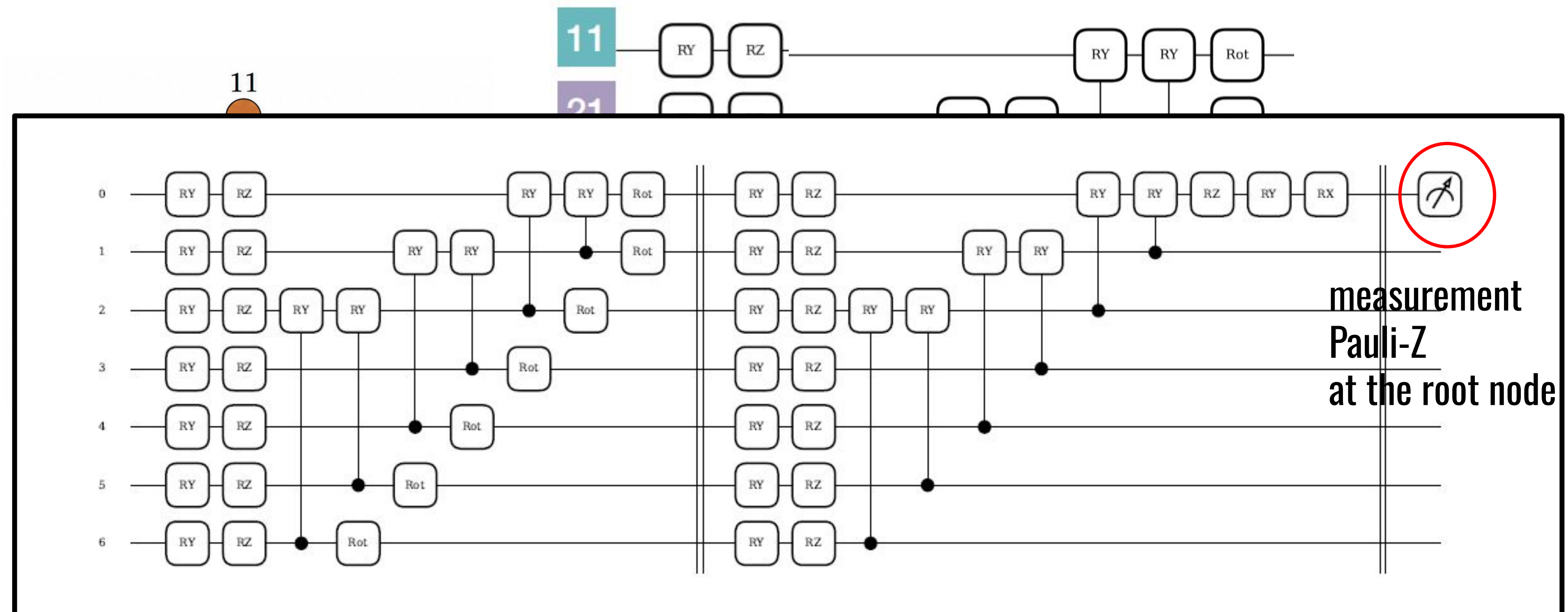
Tree entanglement via controlled CRY gate, node updated with rotations (repeat L-1 times)

$$Rot(\alpha, \beta, \gamma) = R_z(\alpha) R_y(\beta) R_z(\gamma)$$

$$\mathrm{CRY}(w) = \mathrm{CNOT} \cdot \mathrm{RY}(-w/2) \cdot \mathrm{CNOT} \cdot \mathrm{RY}(w/2)$$

Quantum-Tree-Tensor-Network

Building the Quantum ansatz with a tree-tensor-network by reproducing the Lund Tree Graph structure



Tree entanglement via controlled CRY gate, node updated with rotations (repeat L-1 times)

Number of Parameters:

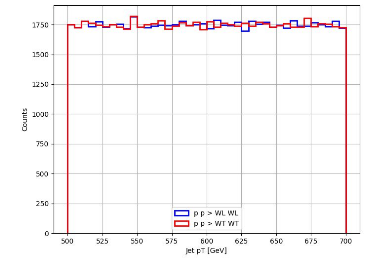
$$|\Theta| = 2N + 6L + 3(L-1)N + 5 = 27L - 2,$$

Datasets

In order to train and test the model we produced different samples, with the aim to be independent from the production, i.e, we want a tagger that take information from the jet structure only and not form high-level variable.

Produced Samples for Polarization Task (MadGraph5+MadSpin+Pythia+Delphes with PU):

1- pp > G > W+W- (Randall Sundrum Model, resonant decay)
2- pp > W+W- (Standard Model, Born level)
3- same-sign Vector Boson Scattering samples in fully hadronic final state (to test the training performed with the previous two samples)



Samples Used for Object Tagging, from JetGame dataset:

1- W vs. Top tagging
2-W vs QCD tagging

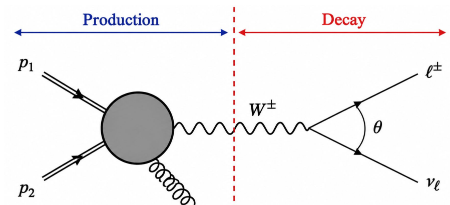
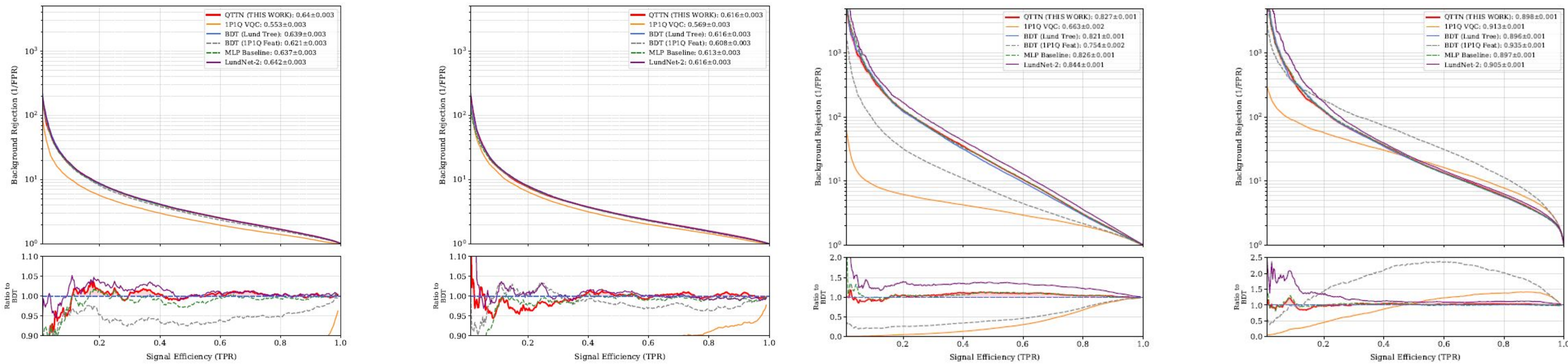


figure adapted from here

Dataset Name	Train Size	Val Size	Test Size	Generator
JetGame W	500,000	50,000	50,000	Pythia
JetGame Top	500,000	50,000	50,000	Pythia
JetGame QCD	500,000	50,000	50,000	Pythia
W^+W^- tree level	196,000	42,000	24,000	Pythia
W^+W^- tree level	98,000	21,000	21,000	Herwig Dipole
W^+W^- tree level	98,000	21,000	21,000	Herwig Angular order
$G_{\text{bulk}} \to W^+W^-$	196,000	42,000	42,000	Pythia

Results: Polarization and Object tagging



(a) Performance on W^+W^- polarization (b) Performance on $G_{\mathrm{bulk}} \to W^+W^-$ polarization (c) Performance on W tagging (d) Performance on Top Tagging

- The QTTN is competitive against the other benchmarks (LundNet, 1P1Q...)
- In the Polarization task, the same results as the modified-LundNet2* (with 3 orders less of parameters)

QTTN (THIS WORK)	1 Layer, 25 params	0.706 ± 0.002
MLP	33 params	0.728 ± 0.002
QTTN	3 Layers, 79 params	0.795 ± 0.001
MLP	79 params	0.806 ± 0.001
QTTN	5 Layers, 133 params	0.813 ± 0.001
MLP	145 params	0.814 ± 0.001
QTTN	10 Layers, 268 params	0.827 ± 0.001
MLP	273 params	0.826 ± 0.001
BDT (QTTN inputs)	13k params	0.821 ± 0.001
LundNet-2	391k params	$\mathbf{0.844} \pm 0.001$
1P1Q	23 params	0.663 ± 0.002
BDT (1P1Q inputs)	13k params	0.754 ± 0.002

W tagging

QTTN (THIS WORK)	1 Layer, 25 params	0.804 ± 0.001
MLP	33 params	0.858 ± 0.001
QTTN	3 Layers, 79 params	0.886 ± 0.001
MLP	79 params	0.883 ± 0.001
QTTN	5 Layers, 133 params	0.893 ± 0.001
MLP	145 params	0.891 ± 0.001
QTTN	10 Layers, 268 params	0.898 ± 0.001
MLP	273 params	0.897 ± 0.001
BDT (QTTN inputs)	13k params	0.896 ± 0.001
LundNet-2	391k params	0.905 ± 0.001
1P1Q	23 params	0.913 ± 0.001
BDT (1P1Q inputs)	13k params	$\mathbf{0.935} \pm 0.001$

top tagging

QTTN (THIS WORK)	1 Layer 25 params	0.616 ± 0.003
MLP	33 params	0.614 ± 0.003
QTTN	3 Layers, 79 params	0.639 ± 0.003
MLP	79 params	0.629 ± 0.003
QTTN	5 Layers, 133 params	0.639 ± 0.003
MLP	145 params	0.635 ± 0.003
QTTN	10 Layers, 268 params	0.640 ± 0.003
MLP	273 params	0.637 ± 0.003
BDT (QTTN inputs)	13k params	0.639 ± 0.003
LundNet-2	391k params	$\mathbf{0.642} \pm 0.003$
1P1Q	23 params	0.553 ± 0.003
BDT (1P1Q inputs)	13k params	0.621 ± 0.003

Polarization (WW)

QTTN (THIS WORK)	1 Layer, 25 params	0.592 ± 0.003
MLP	33 params	0.598 ± 0.003
QTTN	3 Layers, 79 params	0.615 ± 0.003
MLP	79 params	0.608 ± 0.003
QTTN	5 Layers, 133 params	$\mathbf{0.616} \pm 0.003$
MLP	145 params	0.612 ± 0.003
QTTN	10 Layers, 268 params	$\mathbf{0.616} \pm 0.003$
MLP	273 params	0.613 ± 0.003
BDT (QTTN inputs)	13k params	$\mathbf{0.616} \pm 0.003$
LundNet-2	391k params	$\mathbf{0.616} \pm 0.003$
1P1Q	23 params	0.569 ± 0.003
BDT (1P1Q inputs)	13k params	0.608 ± 0.003

Polarization (Graviton)

MLP, BDT with equal or higher number of parameters as additional benchmarks to test quantum approach

*LundNet2 (lnktMin>1, max-depth=3, using (ln(1/z), ln(1/R))

Systematics From the Different Showering Assumption

We investigate the systematics from the showering algorithm adopted by implementing three of the available showering evolution models on the market (pT-ordering, angular ordering, dipole) and taking Pythia8 as reference

Model	Target Domain: Pythia			Target Domain: Herwig			Mean Δ
	Native (P→P)	Trans. (H→P)	$\Delta_{H\to P}$	Native (H→H)	Trans. (P→H)	$\Delta_{P\to H}$	
Herwig Parton Shower: Angular Order							
QTTN (THIS WORK)	0.640	0.637	0.4%	0.609	0.609	**0.1%**	**0.2%**
BDT	0.639	0.636	0.5%	0.609	0.609	**0.1%**	0.3%
MLP	0.634	0.634	**0.0%**	0.609	0.606	0.4%	**0.2%**
1P1Q	0.549	0.556	1.4%	0.551	0.541	1.7%	1.5%
LundNet-2	0.642	**0.639**	0.5%	0.609	0.610	**0.1%**	0.3%
LundNet-5	**0.643**	0.636	1.2%	0.606	**0.611**	0.8%	1.0%
Herwig Parton Shower: Dipole							
QTTN (THIS WORK)	0.640	0.638	0.3%	0.612	**0.614**	0.2%	**0.2%**
BDT	0.639	0.636	0.5%	0.612	0.613	**0.1%**	0.3%
MLP	0.634	0.633	**0.2%**	0.610	0.611	**0.1%**	**0.2%**
1P1Q	0.549	0.557	1.6%	0.554	0.545	1.8%	1.7%
LundNet-2	0.642	0.638	0.6%	**0.613**	0.613	**0.1%**	0.3%
LundNet-5	**0.643**	**0.639**	0.7%	0.610	**0.614**	0.7%	0.7%

$$\Delta_{S\to T} = \frac{|\mathrm{AUC}_{T\to T} - \mathrm{AUC}_{S\to T}|}{\mathrm{AUC}_{T\to T}}$$

Measuring the transfer gap: model assumptions can give large uncertainties*

*Up to 40% for low-level variables [https://scipost.org/10.21468/SciPostPhys.17.6.166]

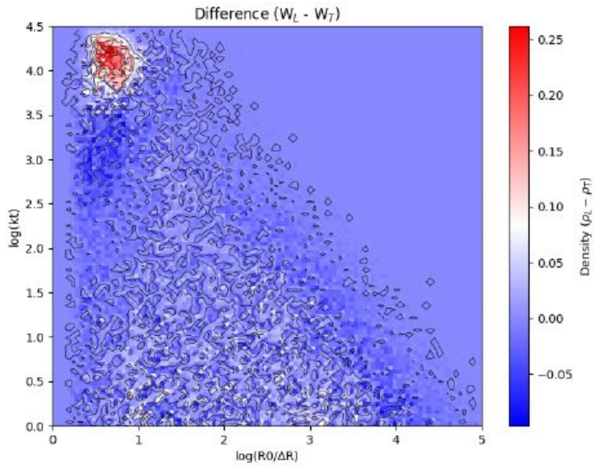
Systematics From the Different Showering Assumption

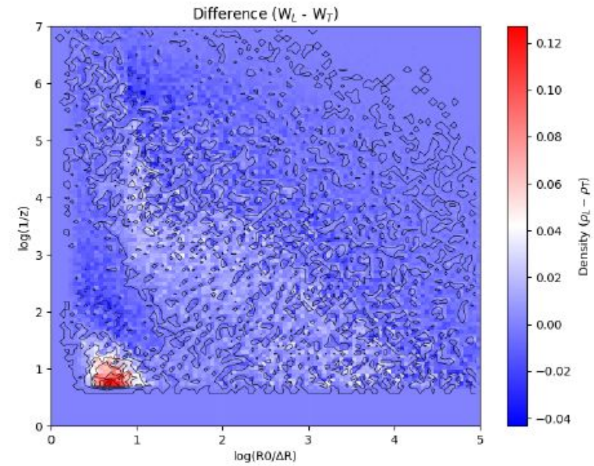
We investigate the systematics from the showering algorithm adopted by implementing three of the available showering evolution models on the market (pT-ordering, angular ordering, dipole) and taking Pythia8 as reference

Model	Target Domain: Pythia			Target Domain: Herwig			Mean Δ
	Native (P→P)	Trans. (H→P)	$\Delta_{H\to P}$	Native (H→H)	Trans. (P→H)	$\Delta_{P\to H}$	
Herwig Parton Shower: Angular Order							
QTTN (THIS WORK)	0.640	0.637	0.4%	0.609	0.609	**0.1%**	**0.2%**
BDT	0.639	0.636	0.5%	0.609	0.609	**0.1%**	0.3%
MLP	0.634	0.634	**0.0%**	0.609	0.606	0.4%	**0.2%**
1P1Q	0.549	0.556	1.4%	0.551	0.541	1.7%	1.5%
LundNet-2	0.642	**0.639**	0.5%	0.609	0.610	**0.1%**	0.3%
LundNet-5	**0.643**	0.636	1.2%	0.606	**0.611**	0.8%	1.0%
Herwig Parton Shower: Dipole							
QTTN (THIS WORK)	0.640	0.638	0.3%	0.612	**0.614**	0.2%	**0.2%**
BDT	0.639	0.636	0.5%	0.612	0.613	**0.1%**	0.3%
MLP	0.634	0.633	**0.2%**	0.610	0.611	**0.1%**	**0.2%**
1P1Q	0.549	0.557	1.6%	0.554	0.545	1.8%	1.7%
LundNet-2	0.642	0.638	0.6%	**0.613**	0.613	**0.1%**	0.3%
LundNet-5	**0.643**	**0.639**	0.7%	0.610	**0.614**	0.7%	0.7%

- MLP, QTTN and LundNet2 have similar transfer gap, compatible with zero
- LundNet5 is always the second with highest transfer-gap

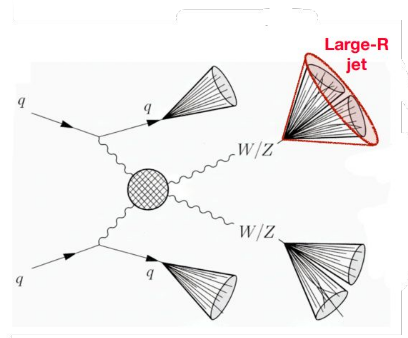
Additional: same-sign VBS fully-hadronic samples



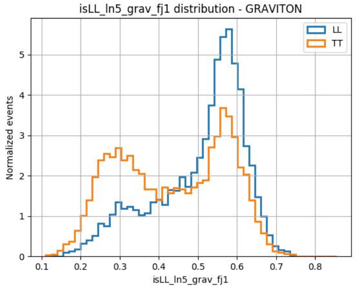


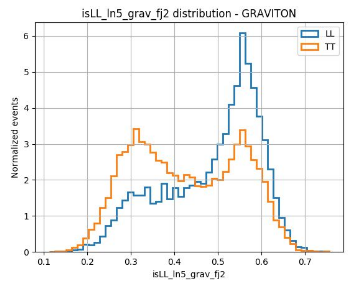
Results with “standard” VBS selection:

- invariant mass of the tagging jets (500 GeV)
- rapidity separation of the tagging jets (2.5)
- exactly 2 AK8



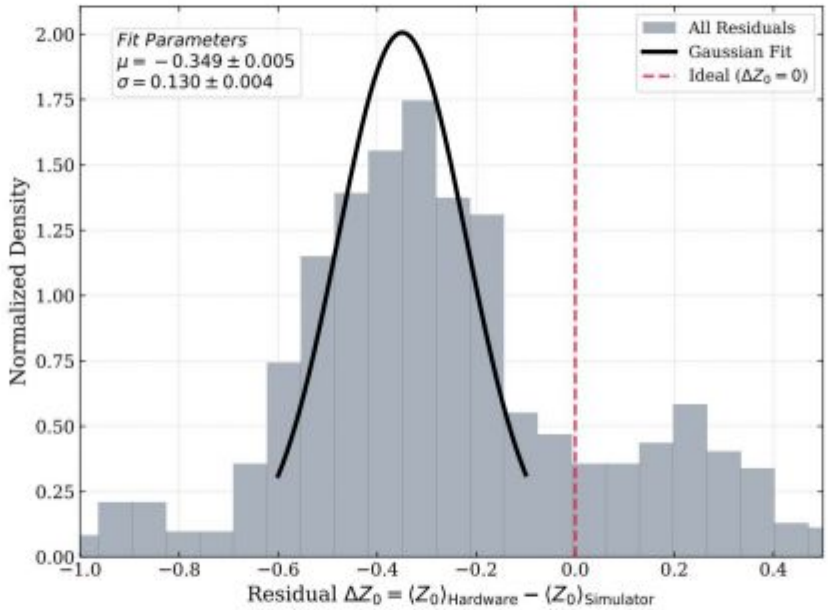
Model	Train Dataset	AUC VBS
LundNet-2	$pp \to W^+W^-$	0.648 ± 0.001
LundNet-5	$pp \to W^+W^-$	**0.664** ± 0.001
QTTN	$pp \to W^+W^-$	0.644 ± 0.002
LundNet-2	RS graviton	0.656 ± 0.001
LundNet-5	RS graviton	0.655 ± 0.001
QTTN	RS graviton	**0.660** ± 0.002





Application on a real hardware

A simplified QTTN version (D=2, L=1) was validated on a real quantum hardware (SpinQ triangulum) with the polarized samples.



AUC (Simulations): 0.60 ± 0.02
AUC (Triangulum): 0.58 ± 0.02

Conclusions

We introduce the L2B encoding with a QTTN, reproducing the LundJet history

-comparative results w.r.t other benchmarks for jet tagging
-test on a real quantum hardware
-inclusion of the polarization task for EW processes
-low parameter count

Thanks to UniNuvola

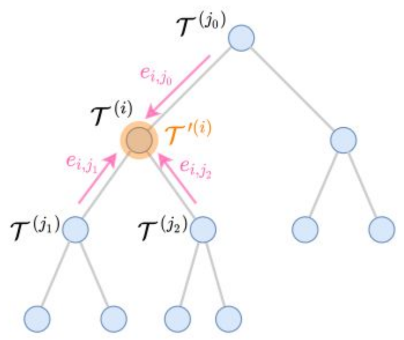
THANKS A LOT FOR YOUR ATTENTION!

BACKUP

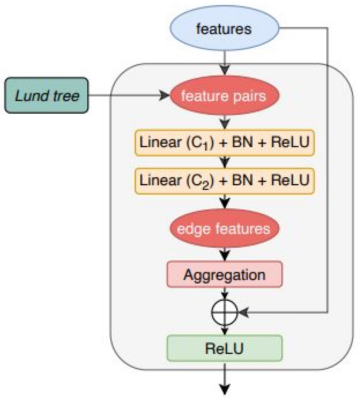
LundNet

A GNN based on the LundPlane is already available: LundNet
Historically: between ParticleNet and ParT, but used only for jet-tagging

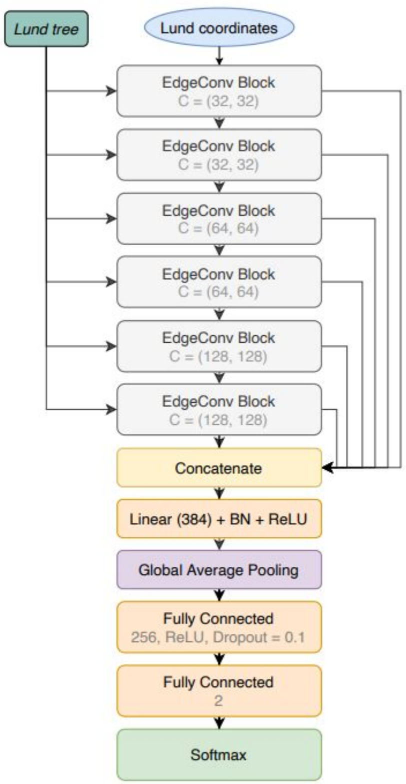
https://arxiv.org/abs/2012.08526



(a)

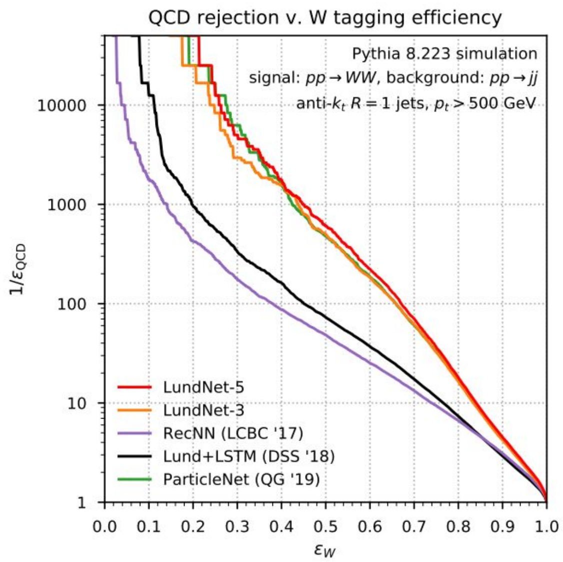


(b)



(c)

$$\mathcal{T}^{(i)} \equiv \{\Delta^{(i)}, k_k^{(i)}, \dots\}$$



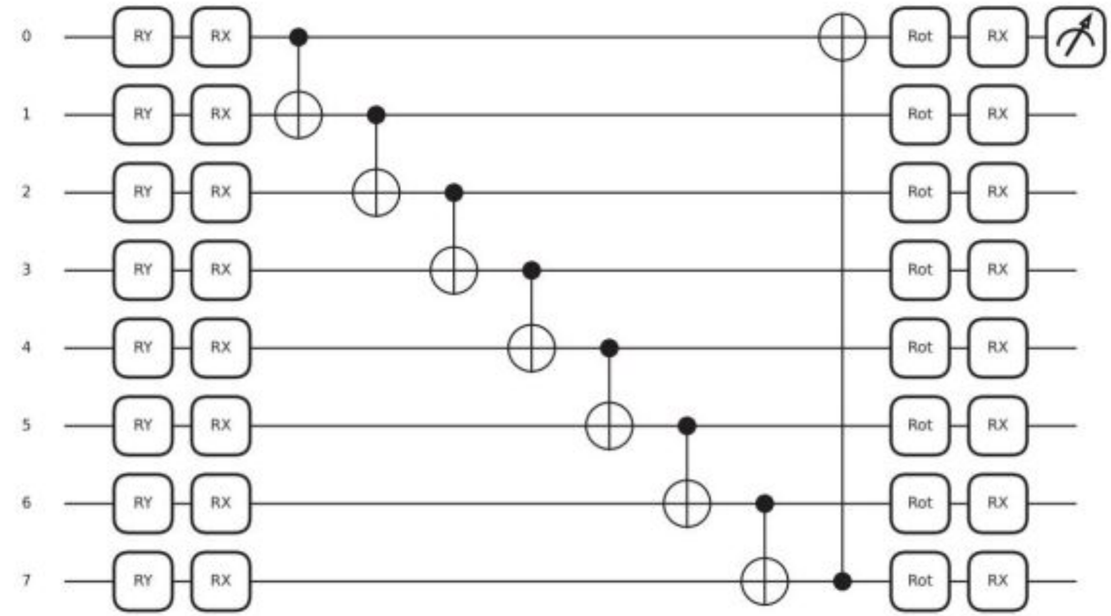
1P1Q

https://journals.aps.org/prd/abstract/10.1103/l8y2-87vq

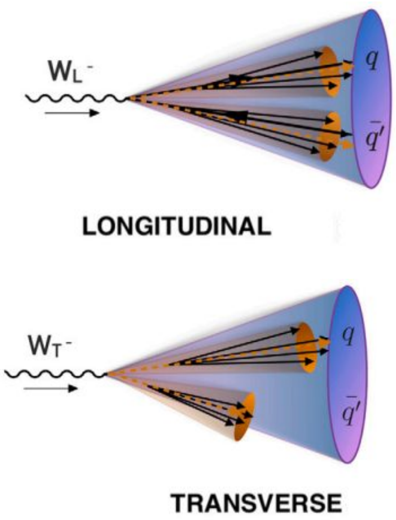
$$f \cdot \frac{p_T}{p_{T,jet}} \cdot (\eta - \eta_{jet}) \rightarrow \theta$$

$$f \cdot \frac{p_T}{p_{T,jet}} \cdot (\phi - \phi_{jet}) \rightarrow \varphi$$

$$f = 1 + 2\pi/(1 + \exp(-w))$$

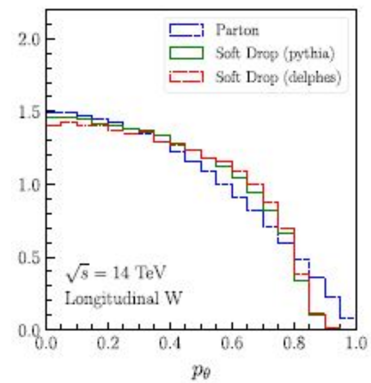


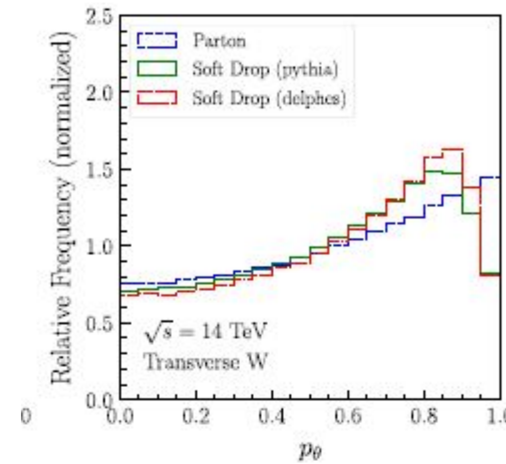
Polarization V-boson related variables

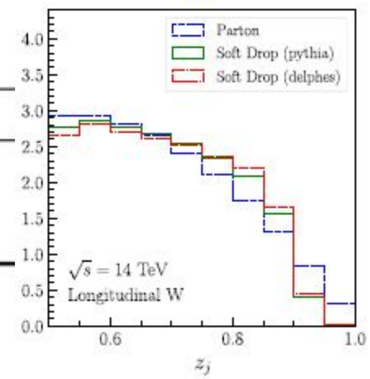


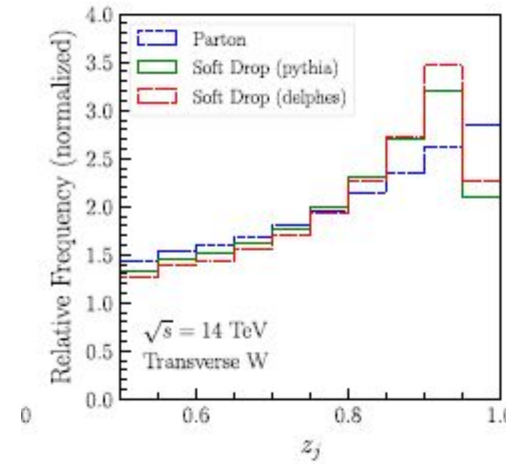
$$p_\theta = \frac{|\Delta E|}{|\vec{p}_W|}$$

$$z_j = \frac{p_T^{\text{leading}}}{p_T^W}$$









Measurement of separability (Area under the ROC curves)

Analysis level and techniques	Longitudinal W Best case scenario	Transverse W Best case scenario
Parton level	0.790	0.789
N-subjettiness, *pythia level*	0.625	0.640
N-subjettiness, *delphes level*	0.610	0.634
Soft Drop, *pythia level*	0.588	0.639
Soft Drop, *delphes level*	0.589	0.632

https://link.springer.com/article/10.1140/epjc/s10052-023-12118-1

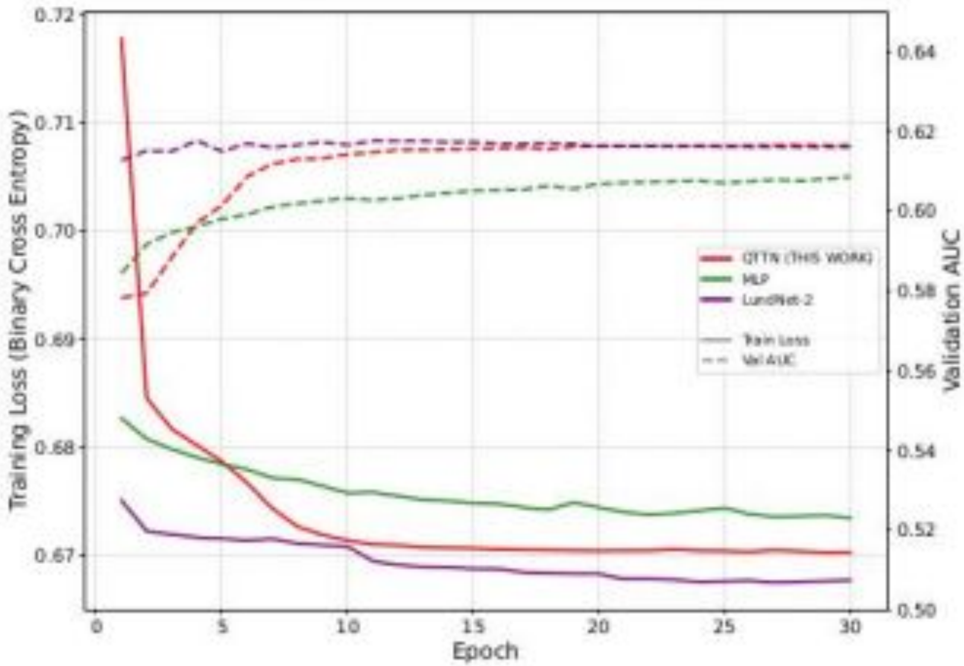


Fig. 5: Training dynamics of the QTTN $L = 3$ for the $G_{\text{bulk}} \to W^+W^-$ process and selected benchmarks over the first 30 epochs, showing the stable minimization of the Binary Cross-Entropy loss alongside the corresponding validation AUC.

PARETO PLOT

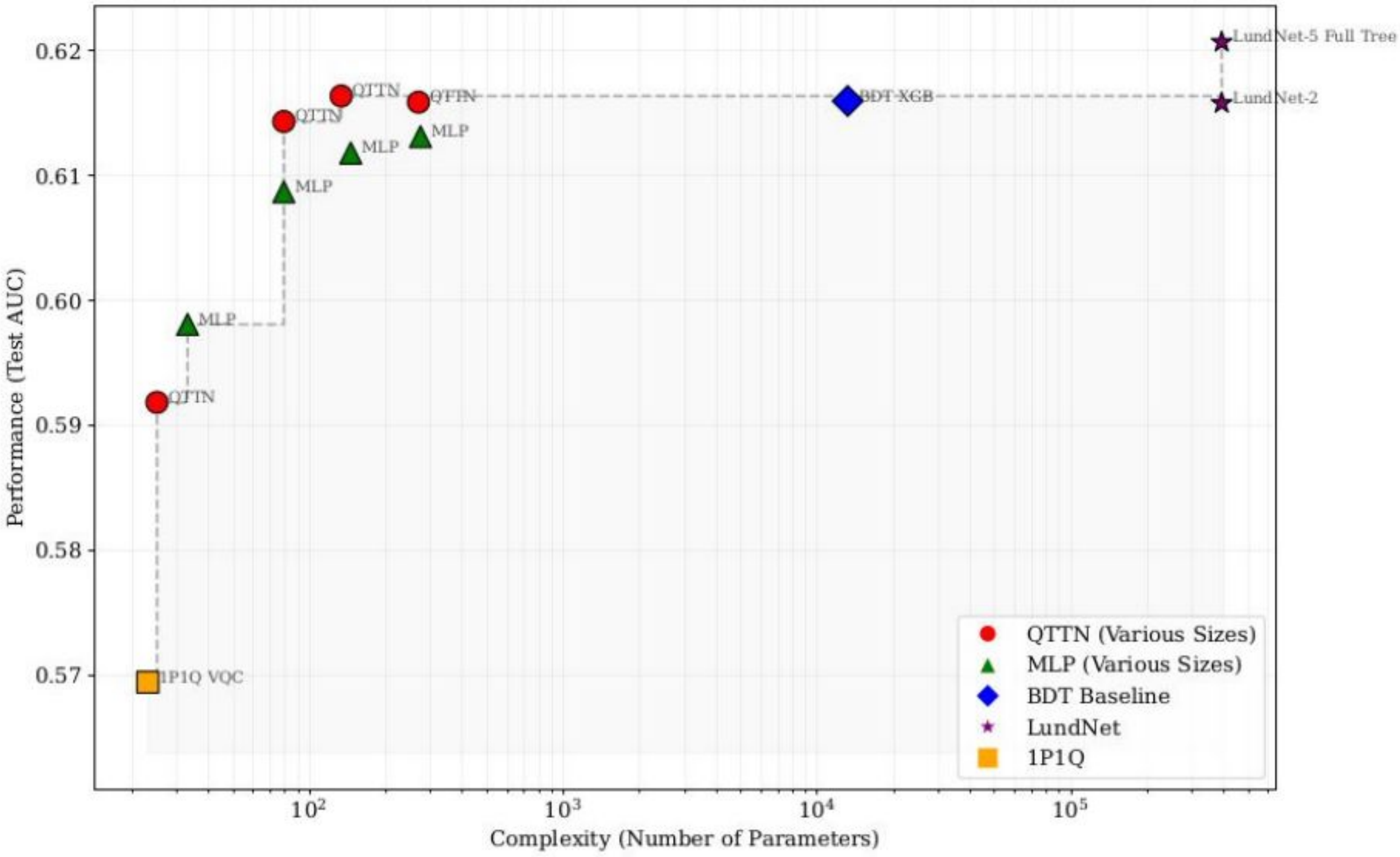
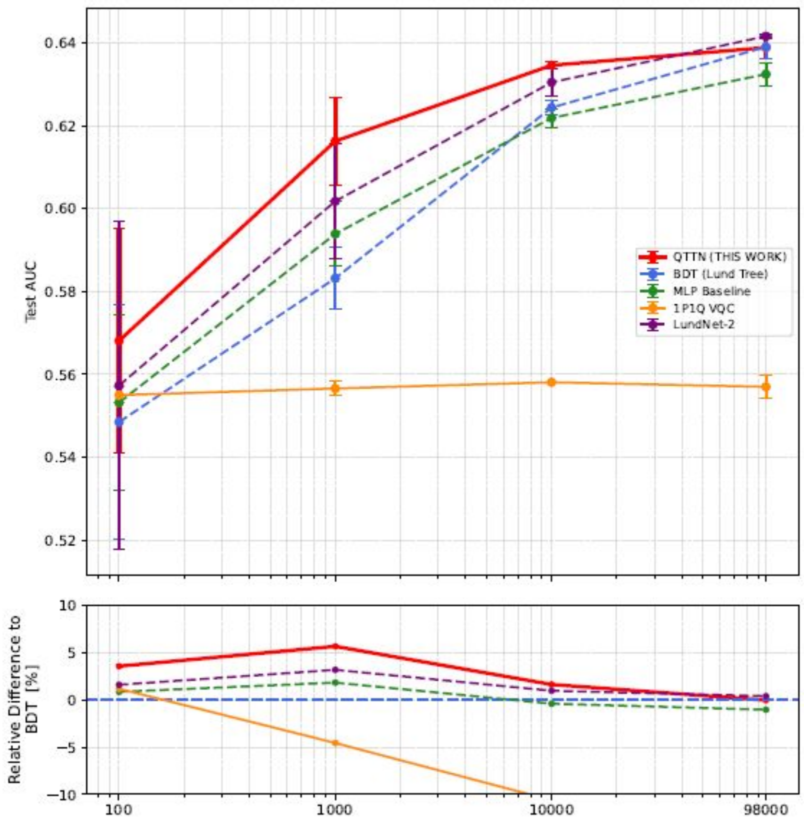
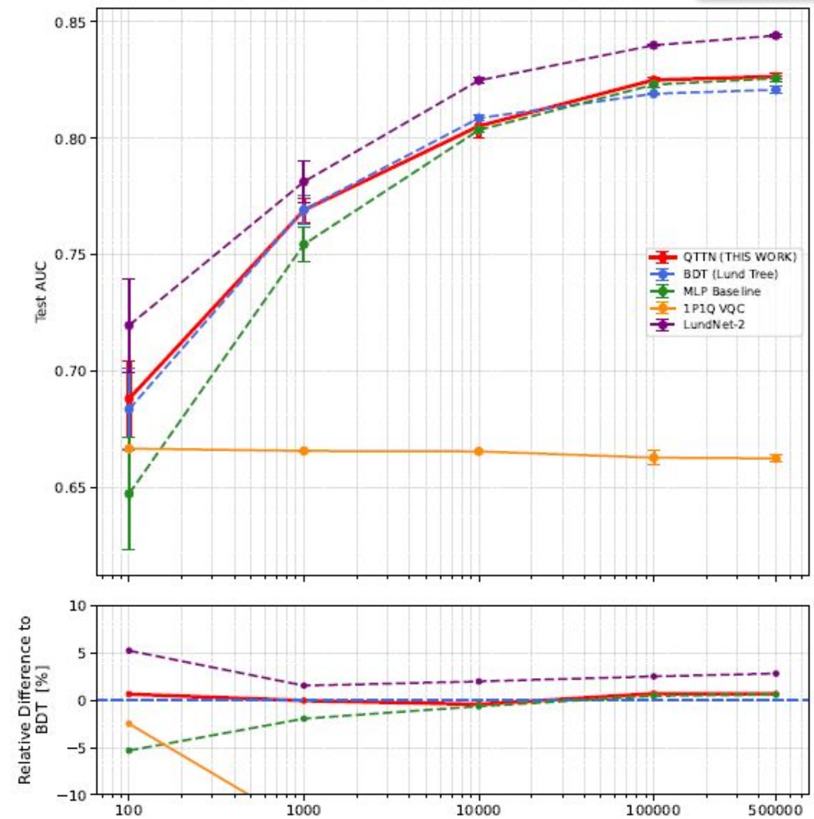


Fig. 7: The Pareto front showing complexity (log scale) versus performance, showing the QTTN (in red) defined in this work with four different layer choices ($L = 1, 3, 5$ and 10), against an MLP baseline of similar parameter count to the QTTN, the BDT trained on QTTN inputs and the LundNet-2 and LundNet-5. QTTN pushes the Pareto front in the low parameter count region. The dataset used is the polarized graviton scenario.

LOW-DATA REGIME

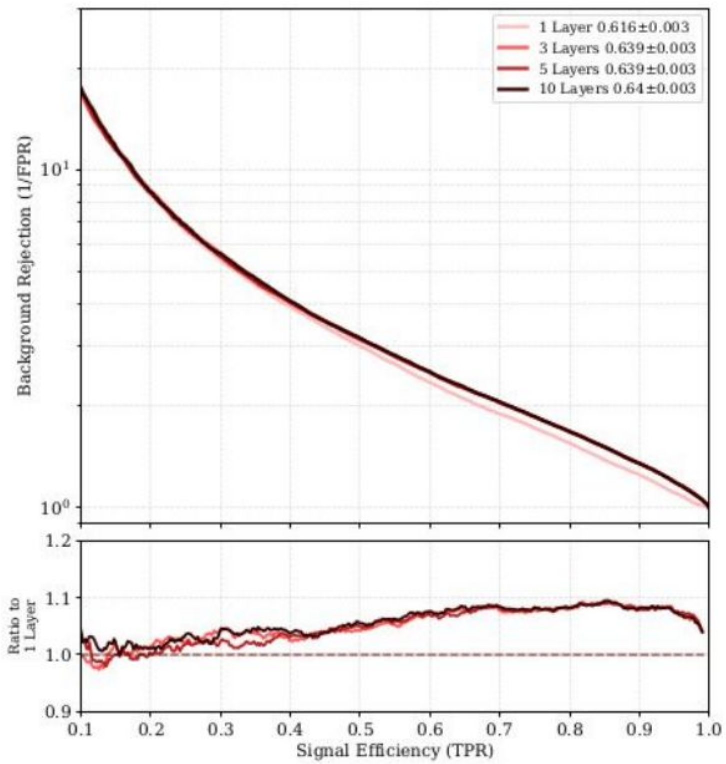


(a) AUC score on the test set as a function of the training data statistics per class, for the W^+W^- tree level sample for the models considered. The red line corresponds to the QTTN with $L = 3$ proposed in this work.

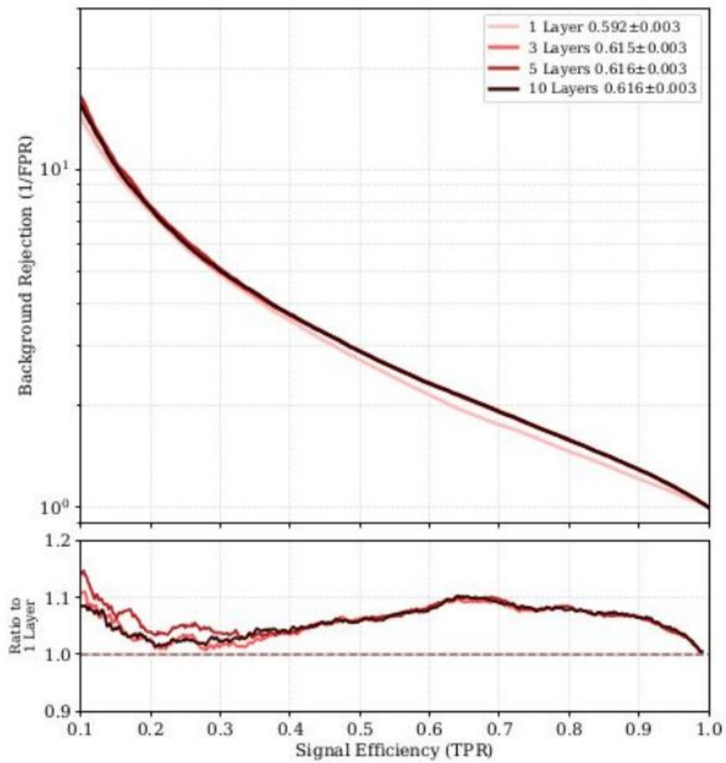


(b) AUC score on the test set as a function of the training data statistics per class, for the W vs QCD sample for the models considered. The red line corresponds to the QTTN with $L = 10$ proposed in this work.

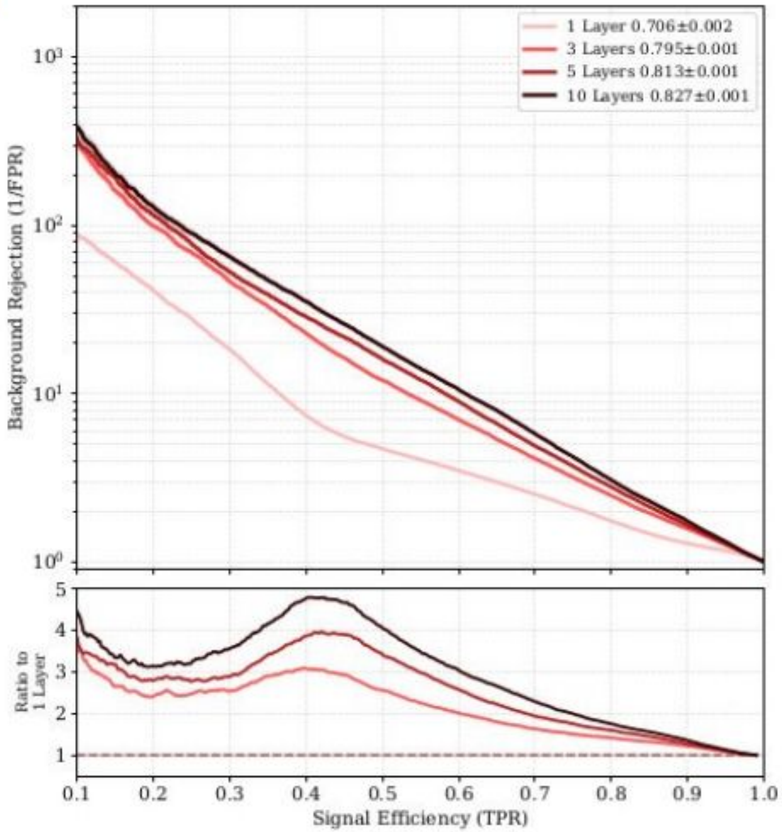
Performance vs Layers



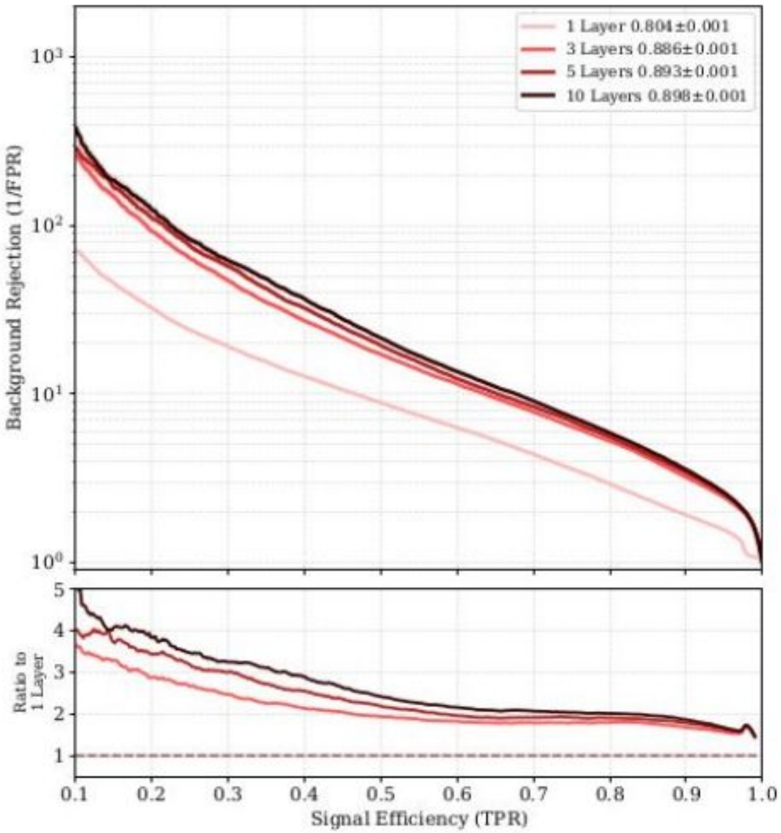
(a) Performance on W^+W^- polarization



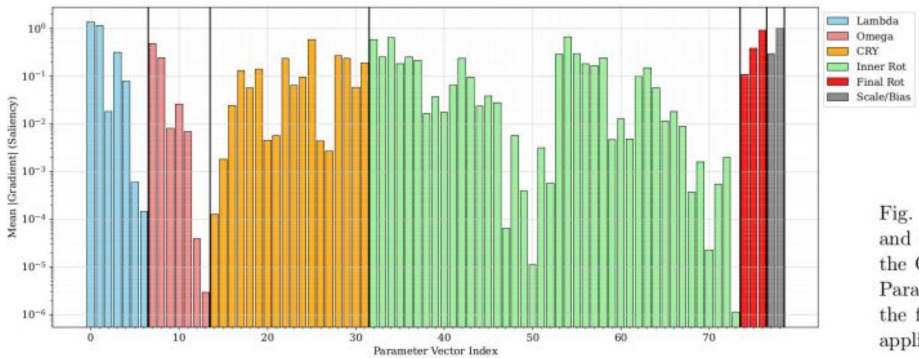
(b) Performance on $G_{\text{bulk}} \rightarrow W^+W^-$ polarization



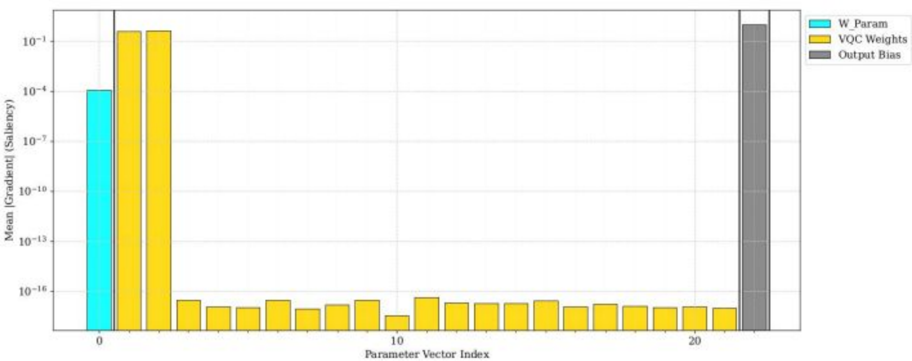
(c) Performance on W tagging



(d) Performance on Top Tagging



(a) Mean saliency for the QTTN quantum circuit.



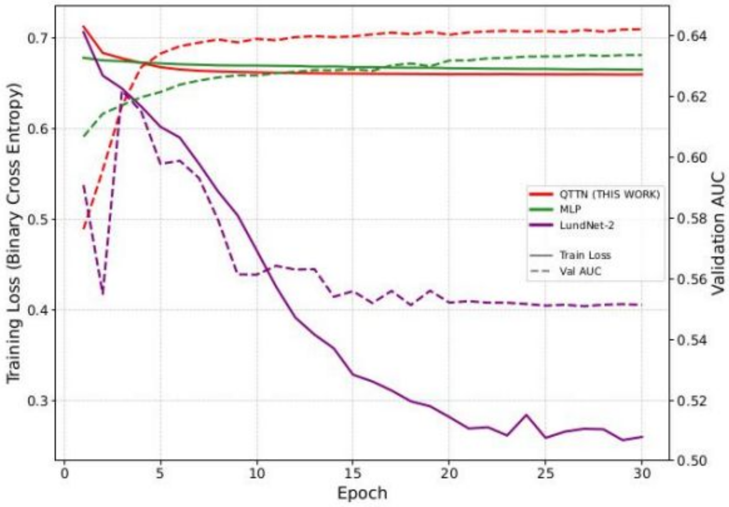
(b) Mean saliency for the 1P1Q quantum circuit.

Fig. 11: Gradient saliency for the parameters of the quantum circuit considered in this work, the QTTN (top) and the 1P1Q (bottom). The color code identifies the different parameter regions in the quantum circuits. For the QTTN, parameters 0-13 are related to the differentiable deformation of the LP2B mapping (blue and red). Parameters 14-31 describe the conditional RY rotations (orange), 32-73 the rotations gates (green), and 74-76 the final rotation before qubit readout. The classical parameters 77-78 (grey) correspond to the scale and bias applied to the quantum measurement. For the 1P1Q the first parameter (cyan) corresponds to the w deformation parameter, 1-21 (yellow) the parametrized rotation, and finally, the last parameter (grey) with the final classical bias applied to the quantum measurement.

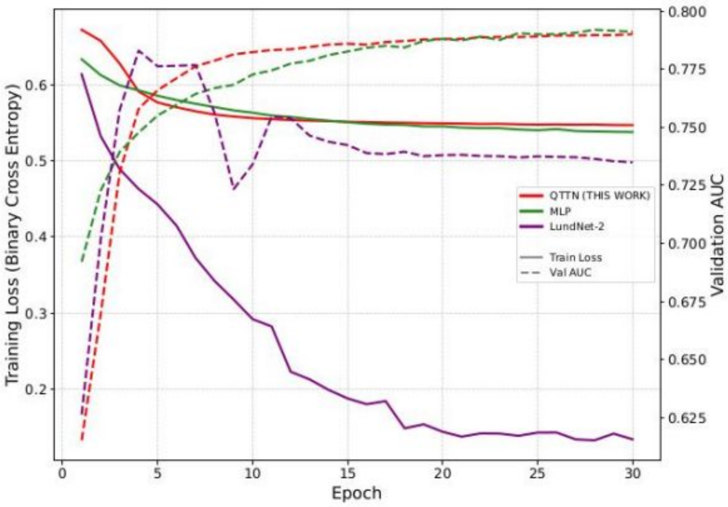
$$S_i = \frac{1}{N} \sum_{j=1}^{N} \left| \frac{\partial y_{\text{logit}}(\boldsymbol{x}_j, \boldsymbol{\Theta}, c_w, c_b)}{\partial \phi_i} \right|$$

PARAMETER SALIENCY

TRAINING DYNAMICS



(a) Training dynamics on the W^+W^- polarization task.



(b) Training dynamics on the W tagging task.

Fig. 12: Training dynamics for the polarization (left) and the W tagging task (right), for LundNet (purple), the QTTN proposed in this work (red) and an MLP of similar parameter count as described in the text (green), showing training loss (left y-axis) and validation AUC (right y-axis), for the low data regime of 1,000 jets per class.

	Model	Configuration	AUC	Dataset
	QTTN (THIS WORK)	1 Layer, 25 params	0.706 ± 0.002	W tagging
	MLP	33 params	0.728 ± 0.002	
	QTTN	3 Layers, 79 params	0.795 ± 0.001	
	MLP	79 params	0.806 ± 0.001	
	QTTN	5 Layers, 133 params	0.813 ± 0.001	
	MLP	145 params	0.814 ± 0.001	
🥈	QTTN	10 Layers, 268 params	0.827 ± 0.001	
	MLP	273 params	0.826 ± 0.001	
🥉	BDT (QTTN inputs)	13k params	0.821 ± 0.001	
🥇	LundNet-2	391k params	$\mathbf{0.844} \pm 0.001$	
	1P1Q	23 params	0.663 ± 0.002	
	BDT (1P1Q inputs)	13k params	0.754 ± 0.002	
	QTTN (THIS WORK)	1 Layer, 25 params	0.804 ± 0.001	Top tagging
	MLP	33 params	0.858 ± 0.001	
	QTTN	3 Layers, 79 params	0.886 ± 0.001	
	MLP	79 params	0.883 ± 0.001	
	QTTN	5 Layers, 133 params	0.893 ± 0.001	
	MLP	145 params	0.891 ± 0.001	
	QTTN	10 Layers, 268 params	0.898 ± 0.001	
	MLP	273 params	0.897 ± 0.001	
	BDT (QTTN inputs)	13k params	0.896 ± 0.001	
🥉	LundNet-2	391k params	0.905 ± 0.001	
🥈	1P1Q	23 params	0.913 ± 0.001	
🥇	BDT (1P1Q inputs)	13k params	$\mathbf{0.935} \pm 0.001$	
	QTTN (THIS WORK)	1 Layer 25 params	0.616 ± 0.003	Polar. W^+W^-
	MLP	33 params	0.614 ± 0.003	
	QTTN	3 Layers, 79 params	0.639 ± 0.003	
	MLP	79 params	0.629 ± 0.003	
	QTTN	5 Layers, 133 params	0.639 ± 0.003	
	MLP	145 params	0.635 ± 0.003	
🥈	QTTN	10 Layers, 268 params	0.640 ± 0.003	
	MLP	273 params	0.637 ± 0.003	
🥉	BDT (QTTN inputs)	13k params	0.639 ± 0.003	
🥇	LundNet-2	391k params	$\mathbf{0.642} \pm 0.003$	
	1P1Q	23 params	0.553 ± 0.003	
	BDT (1P1Q inputs)	13k params	0.621 ± 0.003	
	QTTN (THIS WORK)	1 Layer, 25 params	0.592 ± 0.003	Polar. $G_{\text{bulk}} \to W^+W^-$
	MLP	33 params	0.598 ± 0.003	
	QTTN	3 Layers, 79 params	0.615 ± 0.003	
	MLP	79 params	0.608 ± 0.003	
🥇	QTTN	5 Layers, 133 params	$\mathbf{0.616} \pm 0.003$	
🥉	MLP	145 params	0.612 ± 0.003	
🥇	QTTN	10 Layers, 268 params	$\mathbf{0.616} \pm 0.003$	
🥈	MLP	273 params	0.613 ± 0.003	
🥇	BDT (QTTN inputs)	13k params	$\mathbf{0.616} \pm 0.003$	
🥇	LundNet-2	391k params	$\mathbf{0.616} \pm 0.003$	
	1P1Q	23 params	0.569 ± 0.003	
	BDT (1P1Q inputs)	13k params	0.608 ± 0.003	

W tagging top 3:

🥇 LundNet-2 🥈 QTTN 🥉 BDT

Top tagging top 3:

🥇 BDT(1P1Q) 🥈 1P1Q 🥉 LundNet

Polarization WW tagging top 3:

🥇 LundNet-2 🥈 QTTN 🥉 BDT

Polarization Graviton top 3:

🥇 LundNet-2, QTTN, BDT

🥈 MLP 🥉 MLP