Fast pileup synthesis and transformer-based anomaly detection for long-lived particles

SLAC ATLAS Group Meeting

Mar. 22, 2024

Sam Young, rotation student



ATLAS' rich search program for new physics

bit - wift mode / 2 - def frame		Model	S	ignatu	re ∫	L dt [fb-	Mass limit			Reference	
$ \frac{1}{2} = \frac{1}{2} - 1$		$\tilde{q}\tilde{q}, \tilde{q} \rightarrow q \tilde{\chi}_1^0$	0 e, µ mono-iet	2-6 jets 1-3 jets	$E_{T_{miss}}^{miss}$	139 36.1	<i>q</i> [1x, 8x Degen.] 1.0 1.0 <i>q</i> [1x, 8x Degen.] 0.9	1.85	m(x̂_1)<400 GeV m(∂)-m(2⁰1=5 GeV		
$ \frac{1}{2} =$	đ	$\tilde{g}\tilde{g}, \tilde{g} \rightarrow q \tilde{q} \tilde{\chi}_1^0$	0 e, µ				R	1.15-1.95	2.3 m(t ⁰)=0 GeV		
$ \frac{1}{11} $		$\tilde{g}\tilde{g}, \tilde{g} \rightarrow q\tilde{q}(\ell \ell)\tilde{\chi}^{0}_{1}$	ee, µµ	2 jets	E_T^{miss}	36.1		1.2	2.2 m(\hat{k}_1^0)<600 GeV m(\hat{g})-m(\hat{k}_1^0)=50 GeV	1805.11381	
$\frac{1}{16} + \frac{1}{16} $			SS e, µ	6 jets		139		1.15	m(ĝ)-m(λ ⁰ ₁)=200 GeV	1909.08457	
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $			SS <i>e</i> , <i>µ</i>	6 jets		139		1.25	m(ğ)-m($\tilde{\chi}_{1}^{0}$)=300 GeV	1909.08457	
$\frac{2 r}{k_{1}^{2}, -4 r_{1}^{2}} = \frac{2 r}{k_{1}^{2}, -2 r_{1}^{2}} = 2 $					÷				10 GeV<∆m(δ ₁ , X ⁰ ₁)<20 GeV	2101.12527	
$ \begin{array}{c} f_{11}, f_{12}, wat stress and the properties of nature at higher energies than even to the properties of nature at higher energies than even to the properties of nature at higher energies than even to the properties of nature at higher energies than even to the properties of nature at higher energies than even to the properties of nature at higher energies than even to the properties of nature at higher energies than even to the properties of nature at higher energies than even to the properties of nature at higher energies than even to the properties of nature at higher energies than even to the properties of nature at higher energies than even to the properties of nature at higher energies than even to the properties of nature at higher energies than even to the properties of nature at higher energies than even to the properties of nature at higher energies than even to the properties of the properties of nature at higher energies than even to the properties of the properties $	5	$b_1b_1, b_1 \rightarrow b\chi'_2 \rightarrow bh\chi'_1$			ETniss	139		0.23-1.35	$\Delta m(\tilde{\chi}_{2}^{0}, \tilde{\chi}_{1}^{0}) = 130 \text{ GeV}, m(\tilde{\chi}_{1}^{0}) = 100 \text{ GeV}$ $\Delta m(\tilde{\chi}_{2}^{0}, \tilde{\chi}_{1}^{0}) = 130 \text{ GeV}, m(\tilde{\chi}_{1}^{0}) = 0 \text{ G}$	1908.03122	
$if_{1}, i_{1}, \dots, i_{1}^{n} (t_{1}^{n}, t_{2}^{n}, t_{1}^{n}, t_$	2	$\tilde{t}_1 \tilde{t}_1, \tilde{t}_1 \rightarrow W b \tilde{\chi}_1^0$	1 e.µ	3 jets/1 b	E_T^{miss}	139			m(\tilde{k}_{1}^{0})=500 G	D hysicists at th	he Large Hadron Collider (LHC) in Europe have
$\frac{1}{12}, \frac{1}{14}, \frac{1}$		$\tilde{\imath}_1 \tilde{\imath}_1, \tilde{\imath}_1 \rightarrow c \tilde{\ell}_1^0 / \tilde{c} \tilde{c}, \tilde{c} \rightarrow c \tilde{\ell}_1^0$	0 e. µ	20	E_T^{miss} E_T^{miss}	36.1 139	ē 0.85	1.4	$m(\tilde{x}_{1}^{0})=0 \text{ G}$ $m(\tilde{r}_{1},\tilde{c})-m(\tilde{x}_{1}^{0})=5 \text{ G}$	explored the	properties of nature at higher energies than eve
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $								1.18		before, and they have	ve found something profound: nothing new.
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	1	$\tilde{\chi}_1^{\pm} \tilde{\chi}_2^0$ via WZ	Multiple ℓ/jet: ee, μμ	s ≥ljet	E_T^{miss} E_T^{miss}	139 139	$\frac{\hat{\chi}_{1}^{4}}{\hat{\chi}_{1}^{4}}$ 0.96 $\frac{\hat{\chi}_{1}^{4}}{\hat{\chi}_{1}^{4}}$ 0.205		$m(\tilde{\chi}_1^0)=0$, wino-b $m(\tilde{\chi}_1^0)=6$ GeV, wino-bino	1911.12606	quanta magazine, 2
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $							<i>x</i> [±] ₁ 0.42				
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	direct			s	E_T^{miss}			6			
$ \frac{1}{4} 1$											
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$			2 e. µ	0 jets	ET	139	ž 0.7		$m(\hat{x}_{1}^{0})=0$	1908.08215	
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $		00 0 . 1000									
State β F-hadron, $\beta \rightarrow q \beta_1^{(1)}$ Multiple 56.1 β β -20 Plum Wise -100 <th< td=""><td>ľ</td><td>nn, n→nG/2G</td><td>4 e,μ 0 e,μ</td><td>0 jets ≥ 2 large je</td><td>$E_{T_{miss}}^{L}$ ets E_{T}^{miss}</td><td>139</td><td>Ĥ 0.55</td><td></td><td>$BR(\hat{\chi}_{1}^{0} \rightarrow Z\tilde{G})=1$</td><td>2103.11684</td><td></td></th<>	ľ	nn, n→nG/2G	4 e,μ 0 e,μ	0 jets ≥ 2 large je	$E_{T_{miss}}^{L}$ ets E_{T}^{miss}	139	Ĥ 0.55		$BR(\hat{\chi}_{1}^{0} \rightarrow Z\tilde{G})=1$	2103.11684	
Multiple Set in a finite state		Direct $\tilde{\chi}_1^+ \tilde{\chi}_1^-$ prod., long-lived $\tilde{\chi}_1^\pm$	Disapp. trk	1 jet	$E_T^{\rm miss}$	139	$\tilde{\chi}_{1}^{\pm}$ 0.66 $\tilde{\chi}_{1}^{\pm}$ 0.21		Pure Wino Pure higgsino	ATLAS-CONF-2021-015 ATLAS-CONF-2021-015	
$R_{1}^{2} \ell_{1}^{2} \ell_{1}^{2}$ L_{μ} R_{μ}^{2}	artic	Stable § R-hadron					Ĩġ				
$\begin{array}{c c c c c c c c c c c c c c c c c c c $			Direl Ire	Multiple	emiss			2.05			
$\begin{array}{c} \left(\frac{1}{2},\frac{1}{$	ľ	$ll, l \rightarrow lG$	Dispi, iep		E_T	139	τ, μ 0.7 τ 0.34				
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$											
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$											
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $		$\tilde{g}\tilde{g}, \tilde{g} \rightarrow qq\chi_1^-, \chi_1^- \rightarrow qqq$ $\tilde{g}, \tilde{g} \rightarrow qq\chi_1^-, \chi_1^- \rightarrow qqq$			115						
$h_{11}^{-},h\rightarrow q^{\ell}$ $2 \mu 2 b 85.1 f_1 \qquad \qquad$	1	$\tilde{t}\tilde{t}, \tilde{t} \rightarrow b\tilde{\chi}_{1}^{\pm}, \tilde{\chi}_{1}^{\pm} \rightarrow bbs$		$\geq 4b$			7 Forbidden 0.95				
1μ DV 136 t_1 [10:10< χ_{11} (10:10< χ_{11} (10:10(χ_{11} (10:10(χ_{11} (10:10(χ_{11} (10:10)(10:10)(10:10)(10:10)(10:10)(10:10)(10:10)(10:10)(10:10)(10:10)(10:10)(10:10)(10:10)(10:10)					b		i [qq, bs] 0.42 0.61				
	1	$t_1 t_1, t_1 \rightarrow q \ell$		2 b DV		36.1 136	$\frac{\tilde{t}_1}{\tilde{t}_1}$ [1e-10< λ'_{11} <1e-8, 3e-10< λ'_{21} <3e-9] 1.0				
		$\tilde{\chi}_1^*/\tilde{\chi}_2^0/\tilde{\chi}_1^0, \tilde{\chi}_{1,2}^0 \rightarrow tbs, \tilde{\chi}_1^+ \rightarrow bbs$					λ ₁ ⁸ 0.2-0.32		Pure higgsino		



ATLAS' rich search program for new physics

	TLAS SUSY S										ATLAS Preliminary $\sqrt{s} = 13 \text{ TeV}$
	Model	5	ignatur	e j	Ldt [fb	.)	Mass limit				Reference
s	$\tilde{q}\tilde{q}, \tilde{q} \rightarrow q \tilde{\chi}_1^0$	0 e, µ mono-jet	2-6 jets 1-3 jets	E_T^{miss} E_T^{miss}	139 36.1	q [1x, 8x Degen.] q [8x Degen.]	1.0 0.9		1.85	m($\tilde{\chi}_1^0$)<400 GeV m($\tilde{\chi}_1^0$)=5 GeV	2010.14293 2102.10874
Inclusive Searches	$\bar{g}\bar{g}, \bar{g} \rightarrow q\bar{q}\tilde{\chi}_1^0$	0 e, µ	2-6 jets	$E_T^{\rm miss}$	139	Ř Ř	Forbidden		2.3 1.15-1.95	m(k ⁰ ₁)=0 GeV m(k ⁰ ₁)=1000 GeV	2010.14293 2010.14293
Se	$\bar{g}\bar{g}, \bar{g} \rightarrow q\bar{q}W\tilde{\chi}_{1}^{0}$	1 e,µ	2-6 jets		139	ğ			2.2	m(x10)<600 GeV	2101.01629
8	$\tilde{g}\tilde{g}, \tilde{g} \rightarrow q\tilde{q}(\ell\ell)\tilde{\chi}_1^0$	ee, µµ	2 jets	E_T^{miss}	36.1	Ř		1.2		m(g)-m(x10)=50 GeV	1805.11381
clusiv	$\tilde{g}\tilde{g}, \tilde{g} \rightarrow qqWZ\tilde{\chi}_{1}^{0}$	0 e,μ SS e,μ	7-11 jets 6 jets	$E_T^{\rm miss}$	139 139	ite ite		1.15	1.97	m($\bar{\ell}_1^0$) <600 GeV m(\bar{g})·m($\bar{\ell}_1^0$)=200 GeV	2008.06032 1909.08457
ri Li	$\tilde{g}\tilde{g}, \tilde{g} \rightarrow t \tilde{d} \tilde{\chi}_1^0$	0-1 <i>e</i> ,μ SS <i>e</i> ,μ	3 b 6 jets	$E_T^{\rm miss}$	79.8 139	ğ ğ		1.25	2.25	m($\tilde{\chi}_{1}^{0}$)<200 GeV m($\tilde{\chi}_{1}^{0}$)=300 GeV	ATLAS-CONF-2018-041 1909.08457
	$\tilde{b}_1 \tilde{b}_1$	0 e,µ	2 b	$E_T^{\rm miss}$	139	$\tilde{b}_1 \\ \tilde{b}_1$	0.68	1.255		$m(\tilde{x}_1^0) \le 400 \text{ GeV}$ 10 GeV $\le \Delta m(\tilde{b}_1, \tilde{x}_1^0) \le 20 \text{ GeV}$	2101.12527 2101.12527
S L	$\tilde{b}_1 \tilde{b}_1, \tilde{b}_1 \rightarrow b \tilde{\chi}^0_2 \rightarrow b h \tilde{\chi}^0_1$	0 e.µ	6 b	Emiss	139	b1 Forbidden	0.10.0.05	0.23-1.35	5 Δm(/	$\tilde{t}_{2}^{0}, \tilde{\chi}_{1}^{0}$ = 130 GeV, m($\tilde{\chi}_{1}^{0}$) = 100 GeV	1908.03122

Perhaps we're looking in the wrong spots or for the wrong models? → Need to safeguard against missing new signs of physics

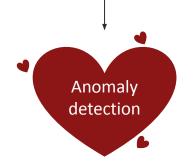
irect $\tilde{\chi}_{1}^{+}\tilde{\chi}_{1}^{-}$ prod., long-lived $\tilde{\chi}_{1}^{\pm}$ table \tilde{g} R-hadron letastable \tilde{g} R-hadron, $\tilde{g} \rightarrow qq \tilde{\chi}_{1}^{0}$ $\tilde{t}, \tilde{\ell} \rightarrow \ell \tilde{G}$	Disapp. trl	Multiple	E_T^{miss}	139	社		0.66				Pure Wino	171 40 0001E 0001 41E
letastable \tilde{g} R-hadron, $\tilde{g} \rightarrow qq \tilde{\chi}_1^0$	8000 PT 1000				A1	0.21					Pure Wino Pure higgsino	ATLAS-CONF-2021-015 ATLAS-CONF-2021-015
	2000 PC 10 PC			36.1	ğ					2.0	I	1902.01636,1808.04095
$l \to \ell \tilde{G}$		Multiple		36.1	ğ [τ(ğ)	=10 ns, 0.2 ns]				2.05 2.4	m(x10)=100 GeV	1710.04901,1808.04095
	Displ. lep		E_T^{miss}	139	ē, µ		0.7				$\tau(\tilde{\ell}) = 0.1 \text{ ns}$	2011.07812
			8		Ŧ	0.34	4				$r(\tilde{\ell}) = 0.1 \text{ ns}$	2011.07812
${}^{*}_{1}\tilde{\chi}^{\mp}_{1}/\tilde{\chi}^{0}_{1}, \tilde{\chi}^{*}_{1} \rightarrow Z\ell \rightarrow \ell\ell\ell$	3 e, µ			139	$\tilde{\chi}_{1}^{+}/\tilde{\chi}_{1}^{0}$	[BR(Zτ)=1, BR(Ze)=1]	0.625	1.05			Pure Wino	2011.10543
${}^{\pm}\tilde{\chi}_{1}^{\mp}/\tilde{\chi}_{2}^{0} \rightarrow WW/Z\ell\ell\ell\ell\nu\nu$	4 e. µ	0 jets	E_T^{miss}	139	$\tilde{\chi}_1^{\pm}/\tilde{\chi}_2^{\pm}$	$[\lambda_{133}\neq 0,\lambda_{12k}\neq 0]$		0.95	1.55		m(\hat{k}_{1}^{0})=200 GeV	2103.11684
$\tilde{g}, \tilde{g} \rightarrow qq \tilde{\chi}_1^0, \tilde{\chi}_1^0 \rightarrow qqq$			ts	36.1)=200 GeV, 1100 GeV]				1.9	Large X''_112	1804.03568
$, \tilde{t} \rightarrow t \tilde{\chi}_{1}^{0}, \tilde{\chi}_{1}^{0} \rightarrow t b s$				36.1	i [X''_323=	2e-4, 1e-2]	0.55					ATLAS-CONF-2018-003
					ī			0.95			m(x11)=500 GeV	2010.01015
$\tilde{t}_1, \tilde{t}_1 \rightarrow bs$		2 jets + 2 b	,		t1 qq. 1	bs]	0.42 0.61					1710.07171
$\tilde{t}_1, \tilde{t}_1 \rightarrow q\ell$	2 e. µ	2 b			ī,	10 - X	-0-0					1710.05544
* *0 *0 · 0					I Line.	104 EV4	c96-91	1.0	1.6		55. 5	2003.11956
$\tilde{\chi}_{1}^{o}/\chi_{2}^{o}/\chi_{1}^{o}, \tilde{\chi}_{1,2}^{o} \rightarrow tbs, \chi_{1}^{o} \rightarrow bbs$	1-2 e, µ	≥6 jets		139	\hat{X}_{1}^{*}	0.2-0.32		_			Pure higgsino	ATLAS-CONF-2021-007
								1 I I				
election of the available mas	ss limits on	new state	s or	1	0^{-1}			1			Mass scale [TeV]	
1 T T T	$[\hat{r}_{1}^{1}\hat{r}_{2}^{1} \rightarrow WWZUtt(v_{Y} \\ \Rightarrow -q_{0}\hat{s}_{1}^{1}, \hat{s}_{1}^{1} \rightarrow qqq$ $\Rightarrow -q \hat{s}_{1}^{1}, \hat{s}_{1}^{1} \rightarrow ts$ $\Rightarrow \delta \hat{r}_{1}, \hat{s}_{1}^{1} \rightarrow bs$ $\hat{r}_{1} \rightarrow bs$ $\hat{r}_{1} \rightarrow te$ $\hat{s}_{1} \rightarrow te$ $\hat{s}_{2}^{1} \hat{s}_{1,2}^{0}, \hat{s}_{1,2}^{1} \rightarrow bbs$ lection of the available main is shown. Many of the interval.	$\left[\frac{1}{2}K_{2}^{2} - \frac{iw_{WZZZUCov}}{2} - \frac{4c_{,\mu}}{2} + \frac{i}{2}K_{2}^{2} + ibs - \frac{i}{2}K_{1}^{2} + ibs - \frac{i}{2}K_{1}^{2} + ibs - \frac{i}{2}K_{1}^{2} + ibs - \frac{i}{2}K_{1}^{2} + \frac{i}{2}bs - \frac{i}{2}K_{1}^{2} + \frac{i}{2$	$ \begin{array}{ll} \left\{ \dot{r}_{1}^{2} \dot{r}_{2}^{2} \rightarrow \psi W Z Z \ell \ell \ell r_{V'} & 4 \ \epsilon_{,\mu} & 0 \ \text{ Jets} \\ -\kappa q \phi q & \xi & \gamma q \phi q \\ -\kappa q \phi q & \xi & \gamma q \phi q \\ -\kappa q & \xi & \gamma + o b s & M \text{ Multiple} \\ -\kappa q & \xi & \xi & \gamma + b s & 2 \ \text{ d} b \\ -\kappa q & \xi & \xi & \gamma + b s \\ -\kappa q & \xi & \chi & \chi \\ -\kappa q & \xi & \chi & \chi \\ -\kappa q & \xi & \chi & \chi \\ -\kappa q & \xi & \chi \\ -\kappa q & \chi & \chi \\ $	$\left[\frac{1}{2}h_{-}^{2} \rightarrow w WZZT(tr_{V} \qquad 4 e, \mu \qquad 0 \text{ [pts } E_{-}^{min} \\ - w WZZT(tr_{V} = q_{0})$ $- w_{1}^{2}, k_{-}^{2} \rightarrow dv$ $- w_{1}^{2}, k_{-}^{2} \rightarrow dv$ $k_{-}^{2}, k_{-}^{2} \rightarrow dv$ $- k_{-}^{2}, k_{-}^{2} \rightarrow$	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$ \begin{array}{c} \frac{1}{2} \frac{1}{2} - \frac{1}{2} - \frac{1}{2} $	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	$ \begin{array}{cccccccccccccccccccccccccccccccccccc$



Reformulating the question

"Does this event look like BSM theory XYZ?"

"Does this event look like the Standard Model?"





Reformulating the question

"Does this event look like BSM theory XYZ?"

Talk focus:Can we correctly identity anomalous events containing long-livedparticles versus regular SM events?





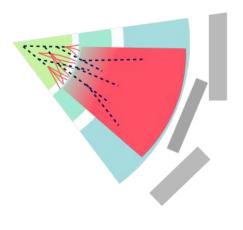
Outline

- Long-lived particles
- Dataset overview
- Fast pileup synthesis
- Event-level classification comparison (MLP v. Transformer)
- Future work



Long-lived particles

- Visible displacement of track or vertex
- >*O*(10) μm decay length (cт)
- DVs not saved by hardware trigger → can we find a way to determine delineate QCD jets from BSM events using low-level information?





Can larger architectures model low-level data well?

- In a single event at the LHC:
 - O(100) vertices
 - *O*(1000) tracks
 - O(10000) hits
- In recent years, more and more complex models like the transformer have been used to model ever-more-complex high-dimensional data to incredible success.
- Goal: moving from high–level jets to low-level tracks, can we adapt these massive models to search for anomalous signals?



Datasets

Signal (LLP)

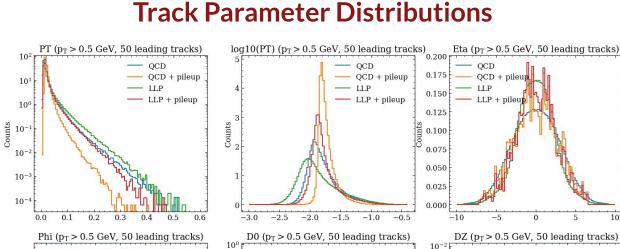
- 200,000 total events
- $p p \rightarrow \tilde{\chi}_3^0 \tilde{\chi}_3^0$
- With pileup (μ =60)
- Two $\tilde{\chi}_3^0$ rest masses: 100 & 500 GeV

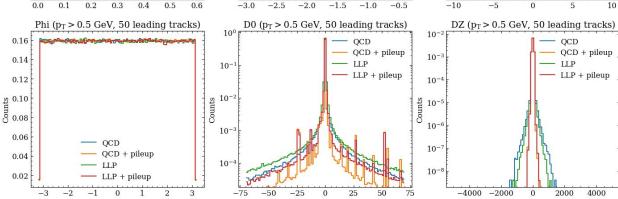
Background (SM)

- 200,000 total events
- $p p \rightarrow 2-5 j$ (pure QCD)
- With pileup (μ=60)

Features

• Each event contains a number of tracks parametrized by $(p_{T}, \eta, \phi, d_{0}, d_{z})$.





q

Datasets

Signal (LLP)

- 200,000 total events
- $p p \rightarrow \tilde{\chi}_3^0 \tilde{\chi}_3^0$
- With pileup (μ=60)
- Two $\tilde{\chi}_3^0$ rest masses: 100 & 500 GeV

Background (SM)

- 200,000 total events
- $p p \rightarrow 2-5 j$ (pure QCD)
- With pileup (μ =60)

Features

• Each event contains a number of tracks parametrized by $(p_T, \eta, \phi, d_0, d_z)$.

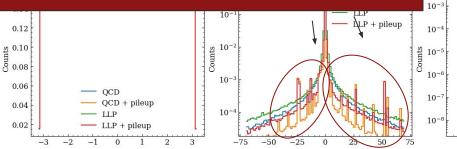
Background Track Parameter Distributions

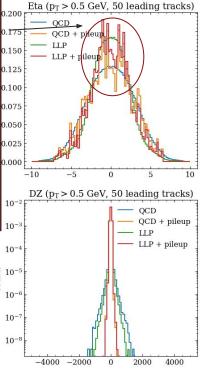
Delphes comes with just 1,000 pileup events to sample from, leading to oversampling by a factor of ~24,000x...

 \rightarrow Need way to simulate ~24 million independent pileup events.

 \rightarrow Simulating pileup is computationally complex. Is there a time- and compute-efficient way to create a synthetic pileup dataset?

 \rightarrow We look into the use of hierarchical gaussian mixture models (HGMMS).







Gaussian Mixture Model (GMM) (normalized)

• To capture event-level correlations, we model each **individual event** track parameter distribution by a weighted mixture of multivariate Gaussians parametrized by means, covariance matrices, and weights:

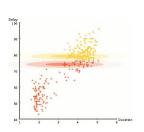
$$p(\vec{x}; \{\mu_i, \Sigma_i, w_i\}_{i=1}^N) = \sum_{i=1}^N w_i \mathcal{N}(\vec{x}; \mu_i, \Sigma_i, w_i), \quad \sum_{i=1}^N w_i = 1.$$

• We use a model selection heuristic Bayesian information criterion (BIC) to choose the number of Gaussians to model each event.

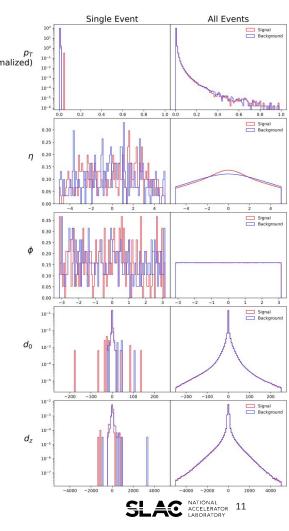
 $BIC = k \log n - 2 \log \hat{L}.$

k is the number of model parameters, n is the number of tracks, and L is the maximized likelihood using the best-fit parameters.

• Balance model complexity with overall fit.

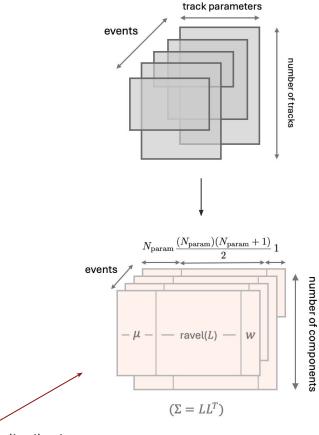


Example of two-component GMM being fit to a two-dim. dataset.



Introducing hierarchy

- After fitting each event's track probability distribution to Gaussian mixtures, we fit a high-complexity Gaussian mixture to the distribution event-level track probability distributions across all events.
- To synthesize new pileup events, we sample from this high-level Gaussian mixture to synthesize a new event-level probability distribution, which is then sampled from to create a variable distribution of particle tracks.



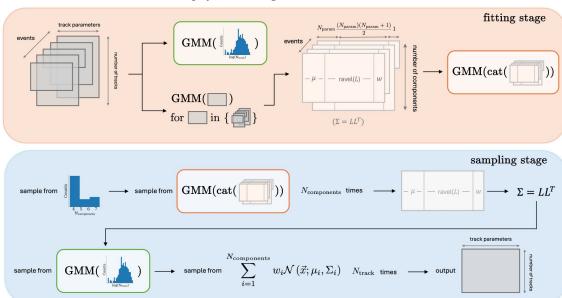
sample new track distributions from a GMM fit to all components



Hierarchical Gaussian Mixture Model (HGMM)

Pros

- Simple idea
- Relatively cheap to sample from compared to actual simulation and non-parametric methods (like KDE)

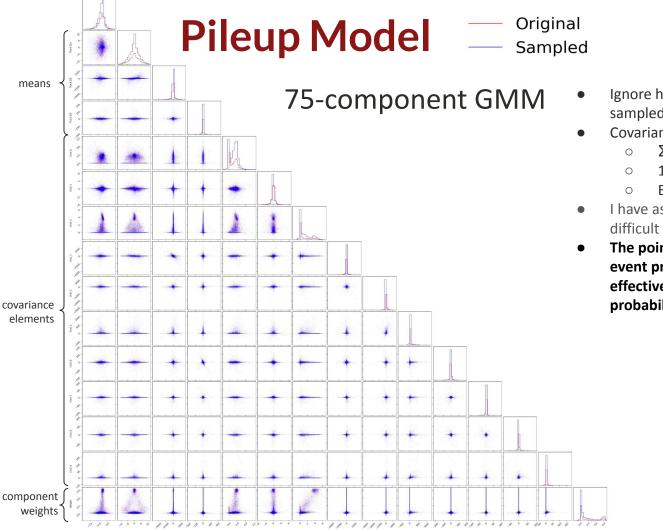


Pileup synthesis using a Hierarchical GMM

Cons

 Assumes track dist's are linear sums of a few multivariate Gaussians (extreme simplification)





Chol_3

Chol.4

Crobs

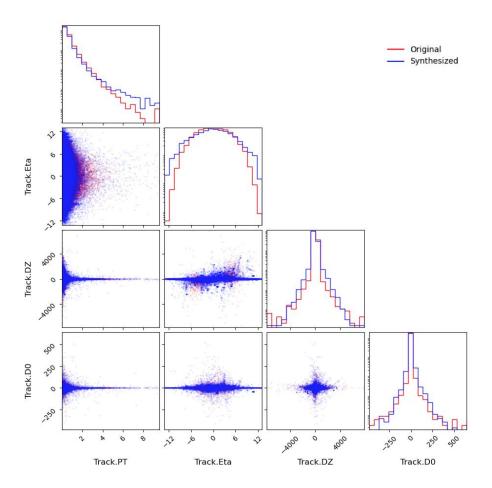
Chol_6

- Ignore histogram scaling (synthetic is doubly sampled)
- Covariance elements are Cholesky decomposed:
 Σ = LL^T
 - \circ 16 elements \rightarrow 10 elements
 - \circ Ensures positive semi-definite nature of Σ
- I have assumed φ is isotropic due to it being difficult to model over 1000 events.
- The point: if we believe that Gaussians model event probability distributions well, we can very effectively model all possible event-level track probability distributions.



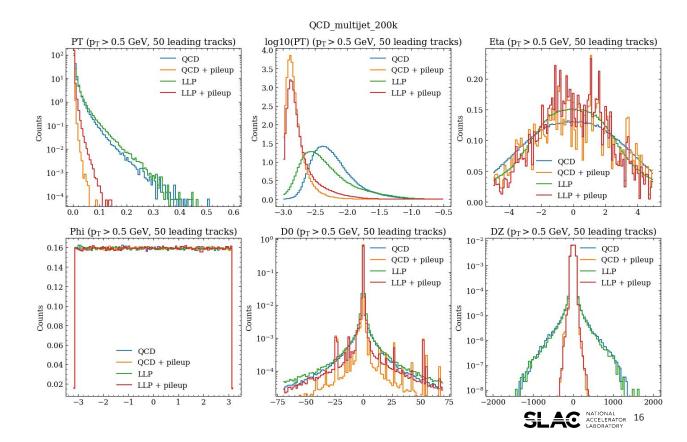
Pileup model (cont.)

- Despite explicitly modeling it, the "global" track probability distributions are well-modeled
- However there are high-covariance "speckles" → overfitting, covariance allowed to be too small for some parameters.
 - Possible fix: scale all parameters to zero mean and unit variance and such that fit covariance matrices have the same scaling between parameters, then clip the covariance to $|\Sigma_{ii}| > \epsilon$
 - Possible fix: fit HGMM to more pileup events.
- Nonetheless, we use this HGMM to synthesize pileup events for our signal/background datasets.



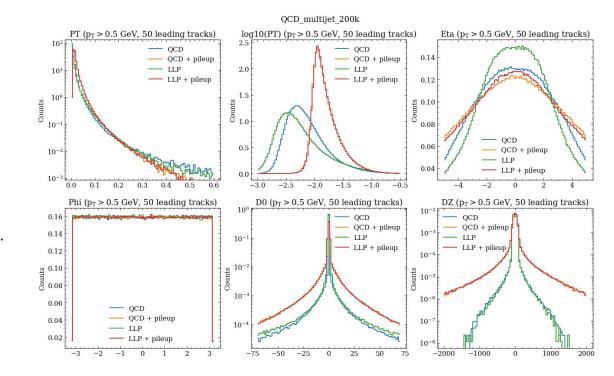


Before: input parameters + oversampled pileup



After: input parameters + synthetic pileup

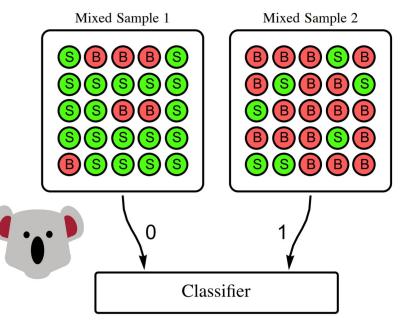
- Issue:
 - Clearly pileup distribution tails aren't well captured (esp. d₀, d₂)
- Possible reasons:
 - We are only fitting 1,000 events and `upscaling' it thousands of times over. Running more events into this model could improve it.





Dataset preprocessing:

- For each training, we apply three cuts:
 - p_τ > 500 MeV
 - |η| < 5
 - \circ Take the 80 tracks with highest p_{T}
- Events are labeled as either 1 (containing LLP) or 0 (pure QCD), meaning that our classifiers are actually being trained on **mixed samples** (a la CWoLA) of tracks.
 - This is what we'd actually see in the LHC, since there's no clear "truth" label anymore when working with tracks.



https://www.ericmetodiev.com/publication/classificationwithoutlabels/



Model architectures:

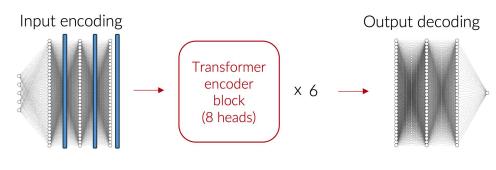
- Goal: Classify an event of tracks as either containing LLP(s) or not.
- We compare two models classifiers:
 - A simple multi-layer perceptron (MLP) with mean reduction along tracks.
 - Transformer-based model
 - MLP encoder (same arch as above)
 - Transformer encoder
 - MLP decoder
- Loss function: binary cross-entropy

$$BCE(X;Y) = \sum_{i=1}^{n} \left\{ y^{(i)} \log x^{(i)} + (1 - y^{(i)}) \log(1 - x^{(i)}) \right\}$$

Transformer-based

MLP

LaverNorm LaverNorm LaverNor

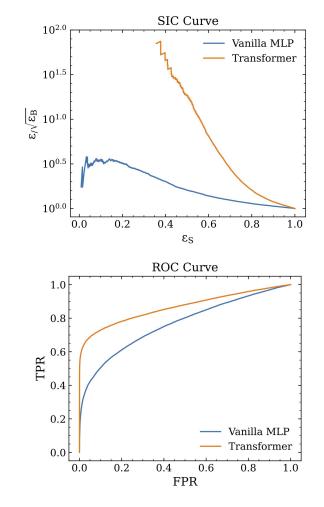




Model performance

→ MLP:

- 70.7% validation accuracy
- 0.768 AUC
- → Transformer:
 - 81.4% validation accuracy
 - 0.860 AUC
- → Side note:
 - Without pileup, classifying between both datasets is trivial (a simple p_T cut gives ~70% accuracy)





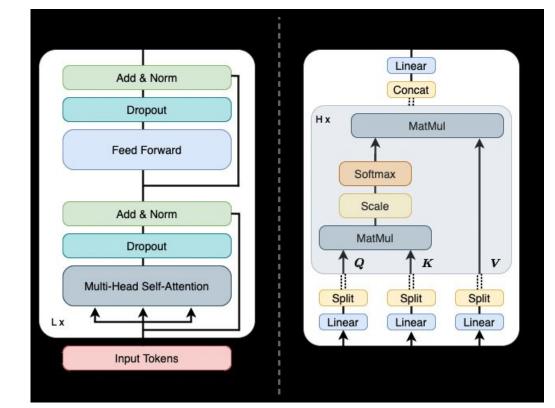
Summary/lessons learned

- Classifying QCD v. LLP events is trivial if there's no pileup
- Modelling pileup using multivariate GMMs is a simple idea but nontrivial in practice.
 - It's better to just simulate the extra events for smaller studies, but as of now simulating huge amount of pileup is hard.
- Transformers can outperform simple MLPs in a high-dimensional task like this.

Backup



Transformer architecture

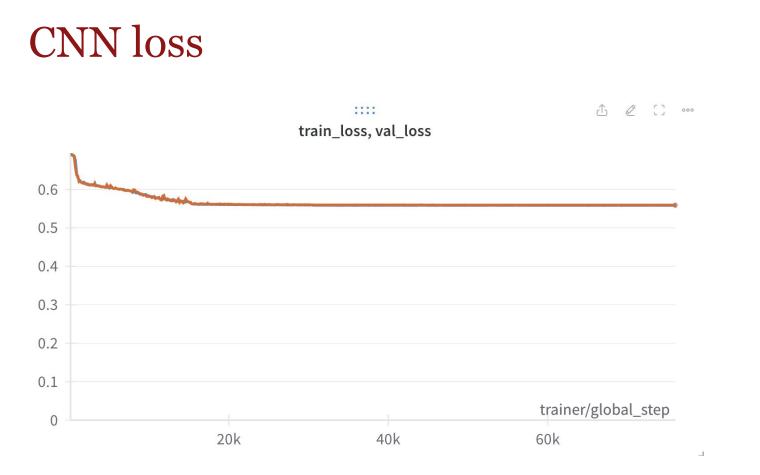


Encoder block

Stanford University



Multi-headed attention





Transformer loss

