

New Approaches to Anomaly Detection

David Shih

February 5, 2020

“Machine Learning at LHC”
KMI Nagoya

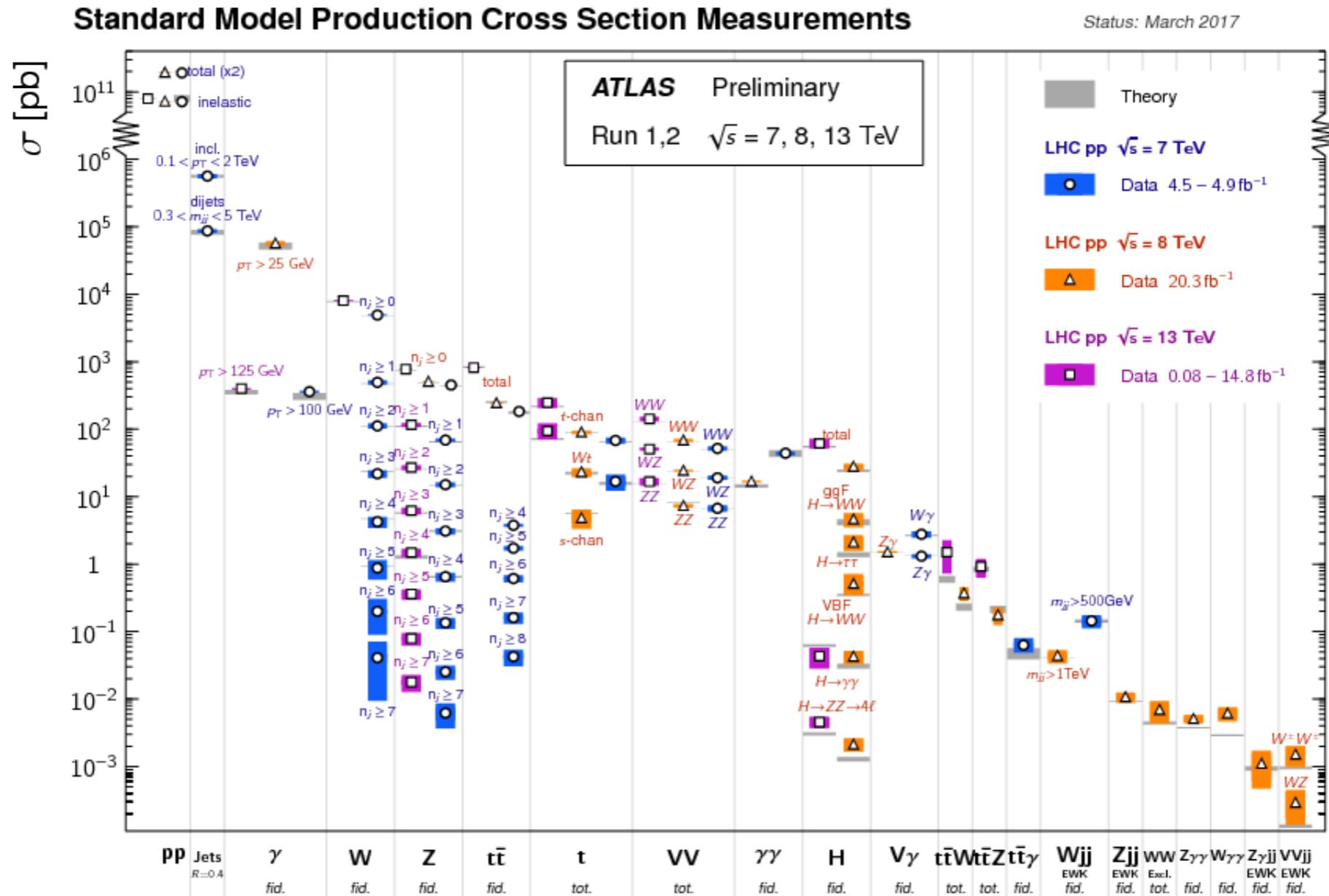


Outline

1. Motivation and setup
2. Conventional approaches
3. New approaches and the LHC Olympics 2020

I. Motivation and Setup

Tests of the Standard Model

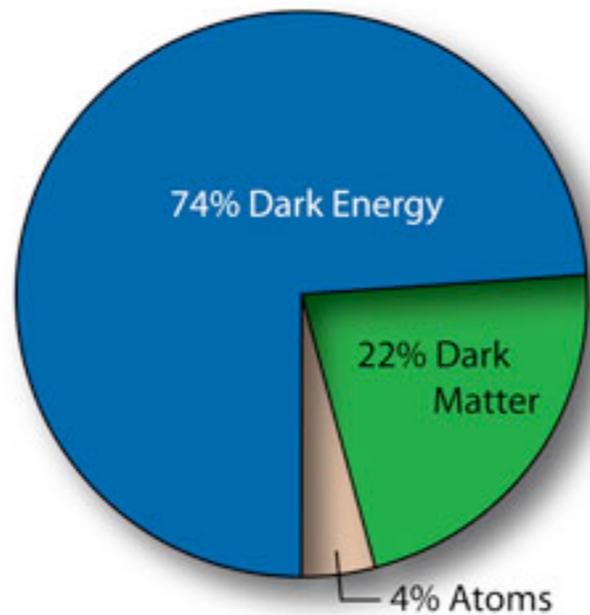


The SM has withstood the test of time. (~50 years!)

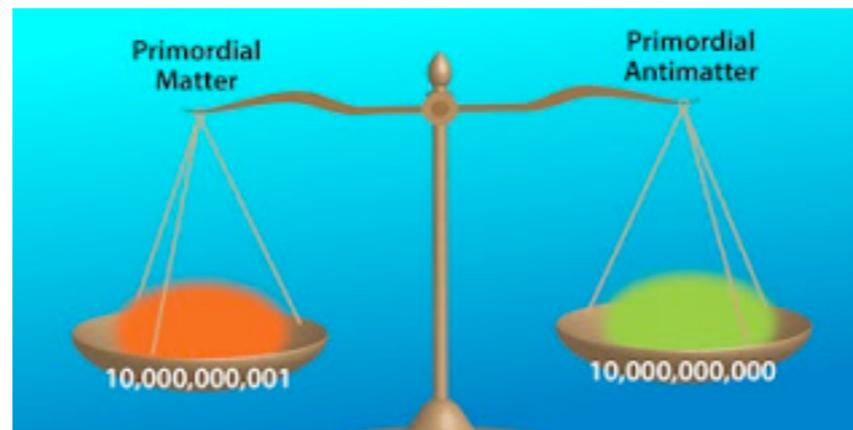
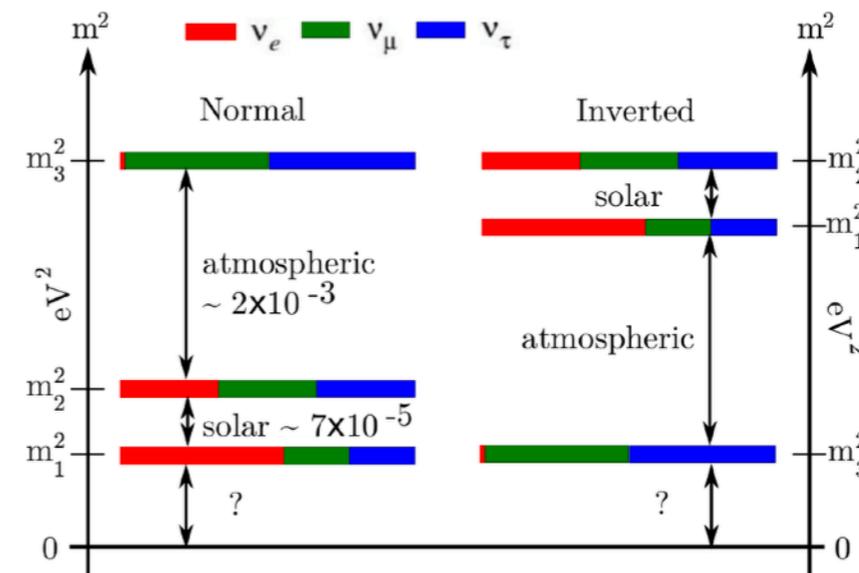
Agreement between theory and experiment across ~14 orders of magnitude.

Yet we know there's new physics out there...

dark matter



neutrino masses

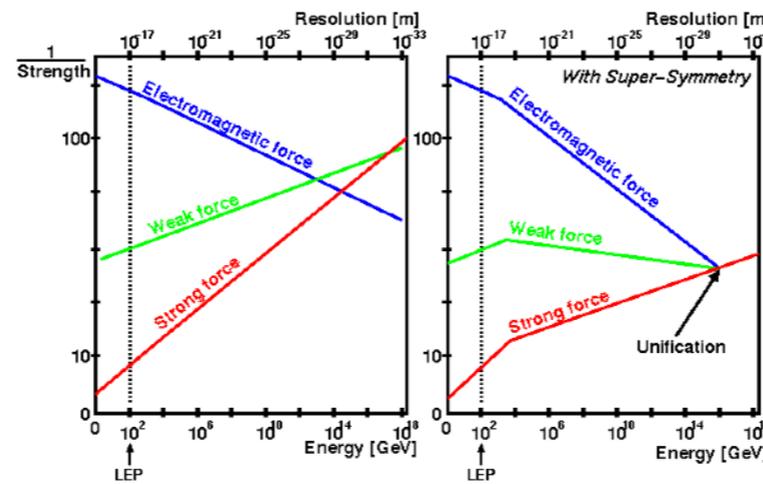
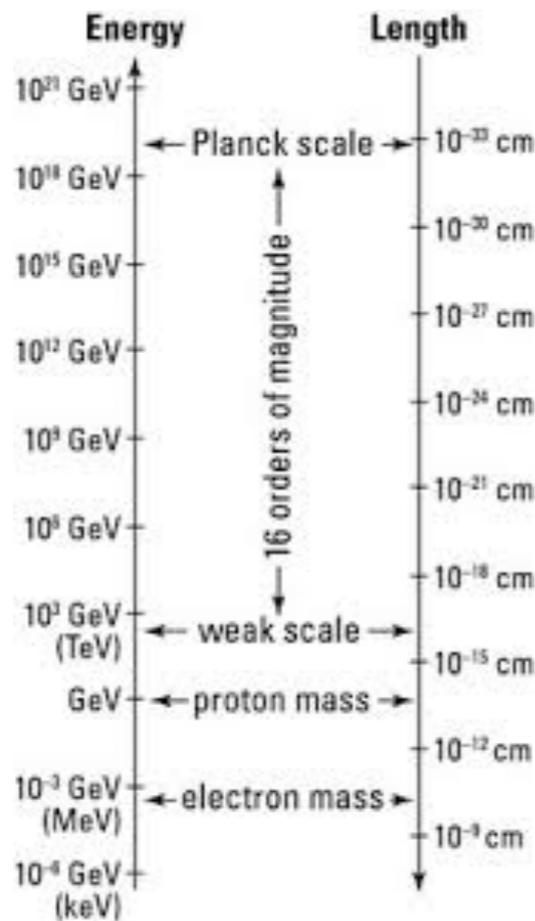


matter/anti-matter asymmetry

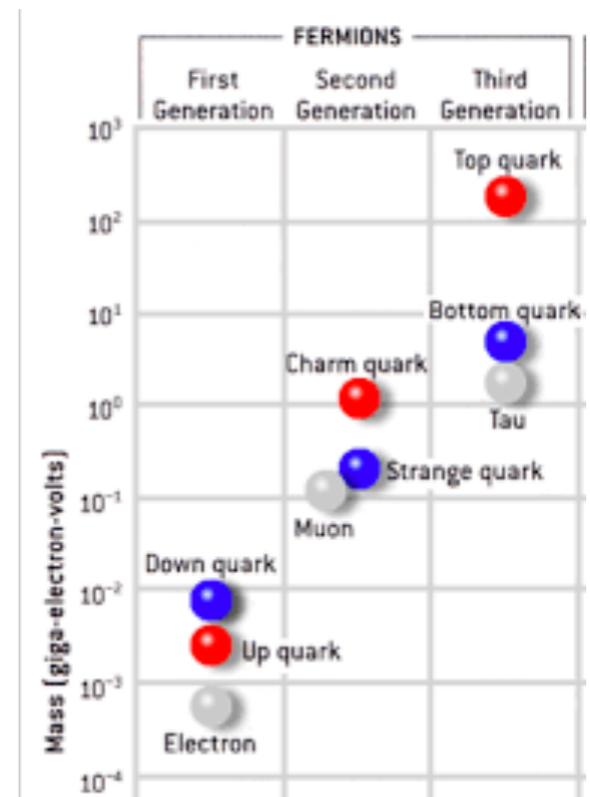
Yet we know there's new physics out there...

grand unification

hierarchy problem



flavor puzzle



$$\mathcal{L} \supset \theta \frac{\alpha_s}{8\pi} G_{\mu\nu} \tilde{G}^{\mu\nu}$$

$$\theta \lesssim 10^{-10}$$

strong CP problem

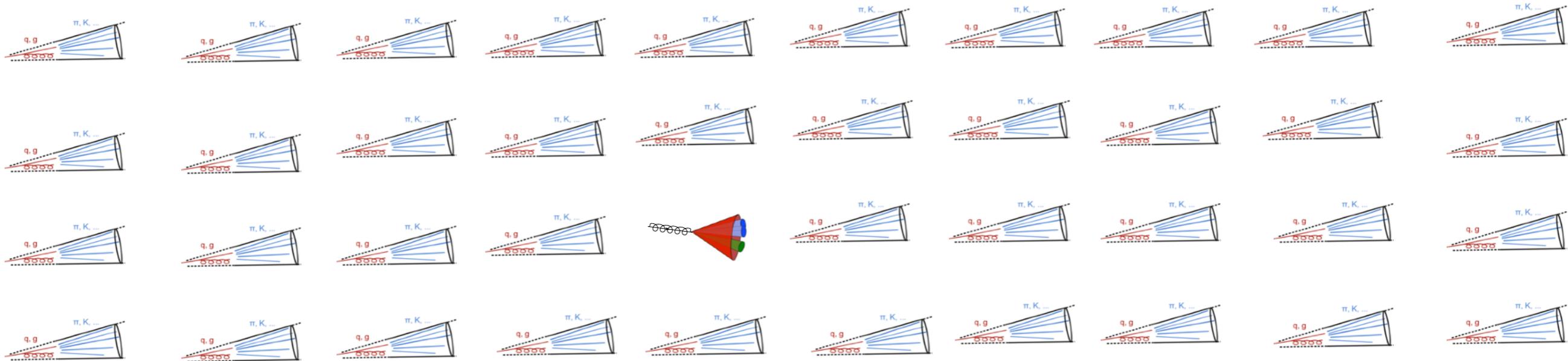
LHC and Big Data

At the LHC we have the highest energy collider ever built, generating copious amounts of data.

- 600 million collisions per second
- Raw data rate ~ 1 PB/s (1 PB = 10^6 GB)
- Actual data rate ~ 25 GB/s
 - Need to trigger on 1 out of 40,000 events
- ~ 10 's of PB annually

***There is enormous discovery potential
for new physics!***

Anomaly Detection at the LHC

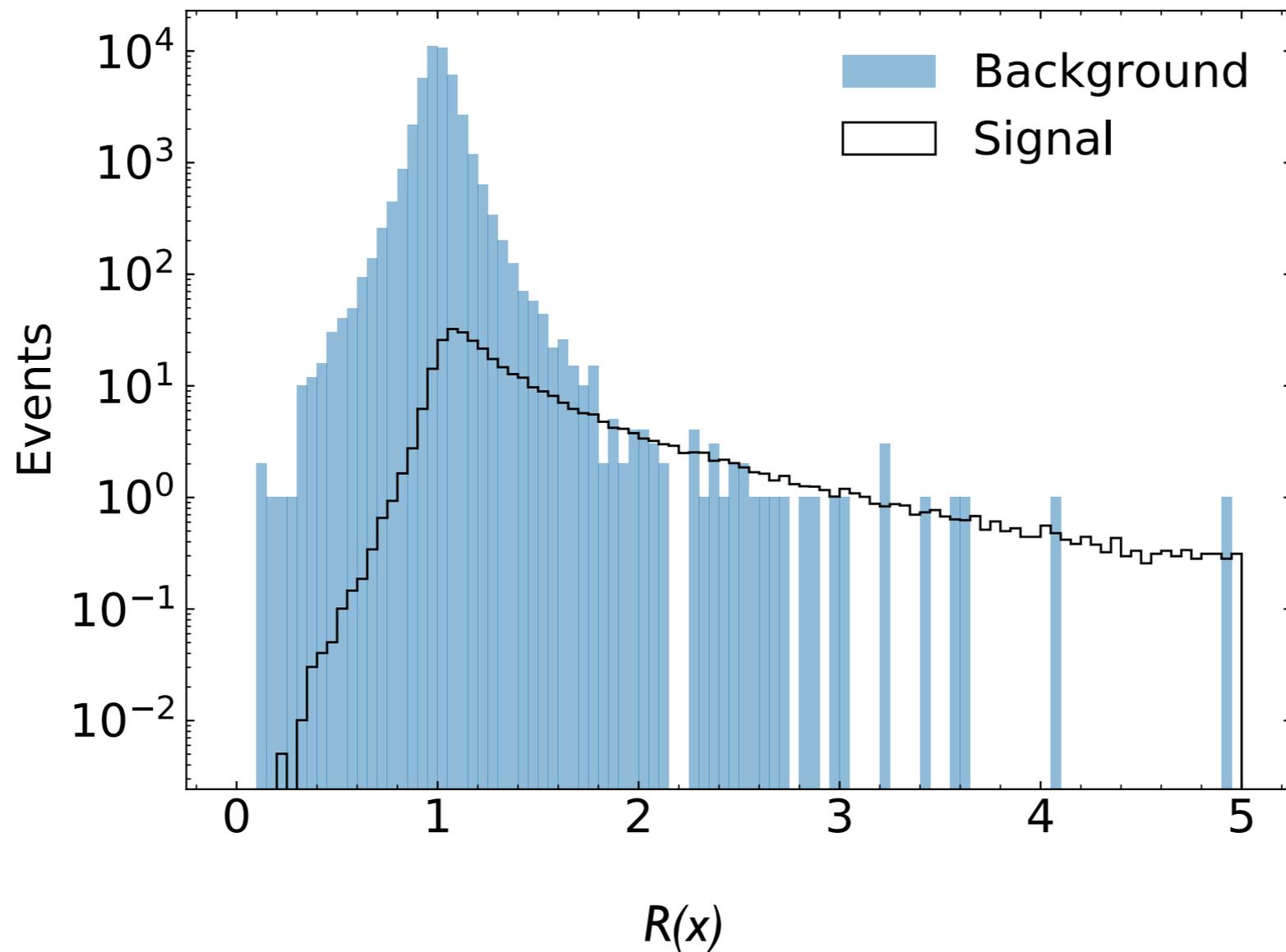


But if there is physics beyond the SM in the data, it's likely to be **rare** and **surrounded by SM background**. *Otherwise we would have seen it already!*

This calls for

- sophisticated techniques to dig the signal out of the data.
- careful and precise background estimation

Anomaly Detection at the LHC

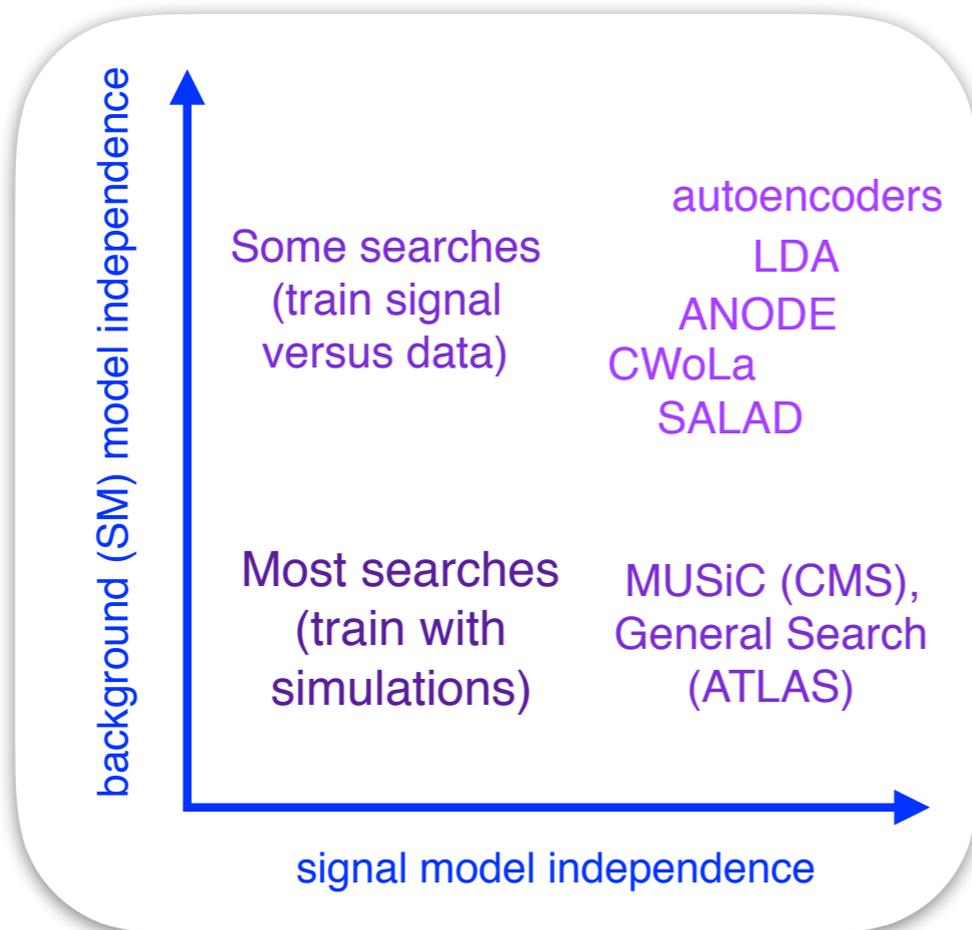


Generally, the idea is to design and optimize a discriminant sensitive to new physics vs. SM background.

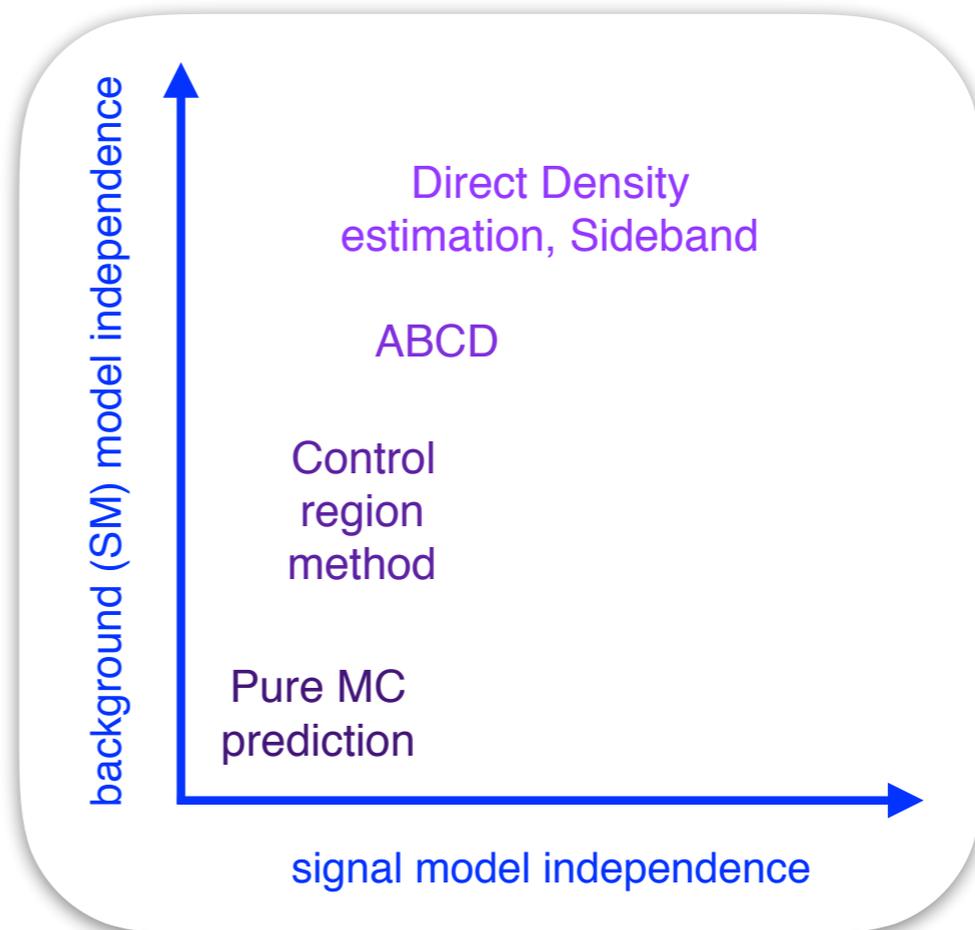
Anomaly Detection at the LHC

Searches make different levels of assumptions about the signal model (e.g. gluinos, general resonance, anything ...) and the background model (SM simulation or data-driven).

from Nachman & DS 2001.04990



(a) Signal sensitivity



(b) Background specificity

2. Conventional methods

The most common approach:

Model specific searches

$$R_S(x) = \frac{\mathcal{L}(x|S_{sim})}{\mathcal{L}(x|B_{sim})}$$

The most common approach:

Model specific searches

$$R_S(x) = \frac{\mathcal{L}(x|S_{sim})}{\mathcal{L}(x|B_{sim})}$$

x: some set of relevant features
characterizing each event (almost
always binned)

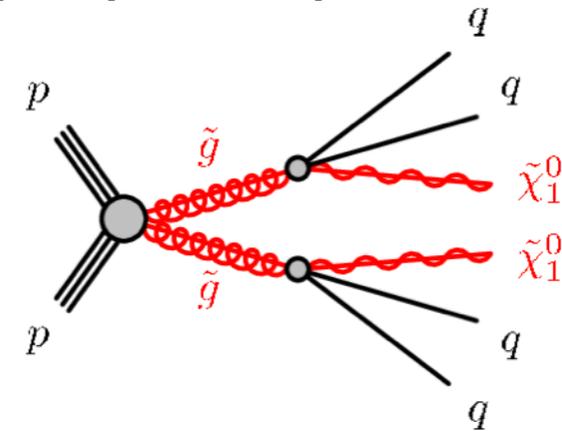
The most common approach:

Model specific searches

$$R_S(x) = \frac{\mathcal{L}(x|S_{sim})}{\mathcal{L}(x|B_{sim})}$$

x : some set of relevant features characterizing each event (almost always binned)

S : a *specific* signal model, e.g. supersymmetry



The most common approach:

Model specific searches

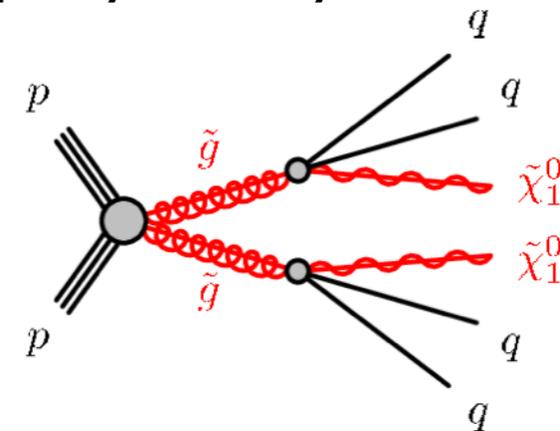
$$R_S(x) = \frac{\mathcal{L}(x|S_{sim})}{\mathcal{L}(x|B_{sim})}$$

x : some set of relevant features characterizing each event (almost always binned)

S : a *specific* signal model, e.g. supersymmetry

Rely on simulations of background and signal to construct likelihood ratio.

If simulations are sufficiently accurate (generally not the case!), then optimal for specific S by Neyman-Pearson Lemma



The most common approach:

Model specific searches

signal and
background model
dependent

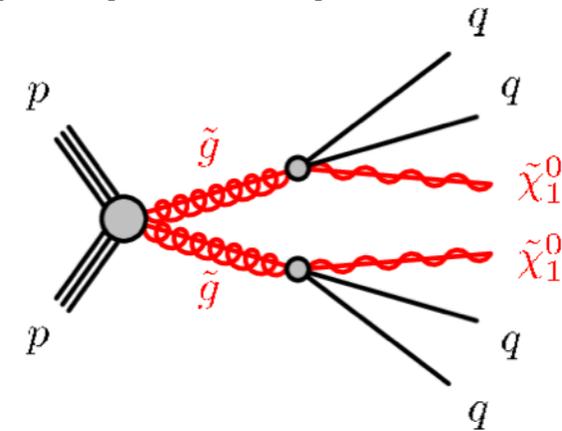
$$R_S(x) = \frac{\mathcal{L}(x|S_{sim})}{\mathcal{L}(x|B_{sim})}$$

x : some set of relevant features characterizing each event (almost always binned)

S : a *specific* signal model, e.g. supersymmetry

Rely on simulations of background and signal to construct likelihood ratio.

If simulations are sufficiently accurate (generally not the case!), then optimal for specific S by Neyman-Pearson Lemma



The most common approach:

Model specific searches

signal and
background model
dependent

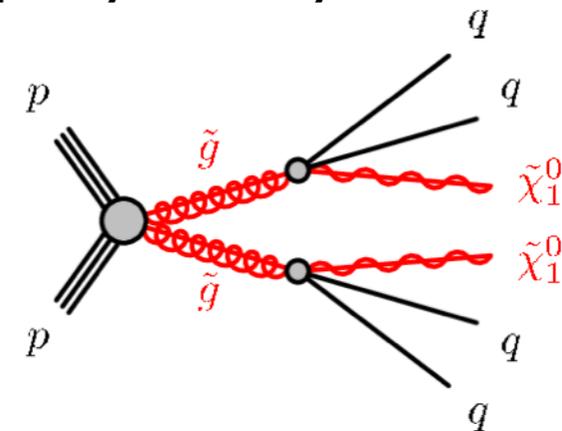
$$R_S(x) = \frac{\mathcal{L}(x|S_{sim})}{\mathcal{L}(x|B_{sim})}$$

x : some set of relevant features characterizing each event (almost always binned)

S : a *specific* signal model, e.g. supersymmetry

Rely on simulations of background and signal to construct likelihood ratio.

If simulations are sufficiently accurate (generally not the case!), then optimal for specific S by Neyman-Pearson Lemma



> 99% of searches at the LHC are of this type

Current Status of NP Searches @ LHC

ATLAS SUSY Searches* - 95% CL Lower Limits

July 2019

ATLAS Preliminary

$\sqrt{s} = 13$ TeV

Model	Signature	$\int \mathcal{L} dt$ [fb ⁻¹]	Mass limit	Reference					
Inclusive Searches	$q\bar{q}, \bar{q} \rightarrow q\tilde{\chi}_1^0$	0 e, μ mono-jet	2-6 jets 1-3 jets	E_T^{miss} E_T^{miss}	36.1 36.1	\tilde{q} [2x, 8x Degen.] \tilde{q} [1x, 8x Degen.]	0.9 1.55	$m(\tilde{\chi}_1^0) < 100$ GeV $m(\tilde{q}) - m(\tilde{\chi}_1^0) = 5$ GeV	1712.02332 1711.03301
	$\tilde{g}\tilde{g}, \tilde{g} \rightarrow q\bar{q}\tilde{\chi}_1^0$	0 e, μ	2-6 jets	E_T^{miss}	36.1	$\tilde{g}\tilde{g}$	2.0	$m(\tilde{\chi}_1^0) < 200$ GeV $m(\tilde{\chi}_1^0) = 900$ GeV	1712.02332 1712.02332
	$\tilde{g}\tilde{g}, \tilde{g} \rightarrow q\bar{q}(\ell\ell)\tilde{\chi}_1^0$	3 e, μ $ee, \mu\mu$	4 jets 2 jets	E_T^{miss} E_T^{miss}	36.1 36.1	$\tilde{g}\tilde{g}$	1.85 1.2	$m(\tilde{\chi}_1^0) < 800$ GeV $m(\tilde{g}) - m(\tilde{\chi}_1^0) = 50$ GeV	1706.03731 1805.11381
	$\tilde{g}\tilde{g}, \tilde{g} \rightarrow qqWZ\tilde{\chi}_1^0$	0 e, μ SS e, μ	7-11 jets 6 jets	E_T^{miss}	36.1 139	$\tilde{g}\tilde{g}$	1.15	$m(\tilde{\chi}_1^0) < 400$ GeV $m(\tilde{g}) - m(\tilde{\chi}_1^0) = 200$ GeV	1708.02794 ATLAS-CONF-2019-015
	$\tilde{g}\tilde{g}, \tilde{g} \rightarrow \tau\tilde{\chi}_1^0$	0-1 e, μ SS e, μ	3 b 6 jets	E_T^{miss}	79.8 139	$\tilde{g}\tilde{g}$	1.25	$m(\tilde{\chi}_1^0) < 200$ GeV $m(\tilde{g}) - m(\tilde{\chi}_1^0) = 300$ GeV	ATLAS-CONF-2018-041 ATLAS-CONF-2019-015
3 rd gen. squarks direct production	$\tilde{b}_1\tilde{b}_1, \tilde{b}_1 \rightarrow b\tilde{\chi}_1^0 / t\tilde{\chi}_1^\pm$	Multiple	Multiple	E_T^{miss}	36.1	\tilde{b}_1	0.9	$m(\tilde{\chi}_1^0) = 300$ GeV, BR($b\tilde{\chi}_1^0$) = 1	1708.09266, 1711.03301
	$\tilde{b}_1\tilde{b}_1, \tilde{b}_1 \rightarrow b\tilde{\chi}_2^0 \rightarrow bh\tilde{\chi}_1^0$	0 e, μ	6 b	E_T^{miss}	139	\tilde{b}_1	0.58-0.82 0.74	$m(\tilde{\chi}_1^0) = 300$ GeV, BR($b\tilde{\chi}_1^0$) = BR($t\tilde{\chi}_1^\pm$) = 0.5 $m(\tilde{\chi}_1^0) = 200$ GeV, $m(\tilde{\chi}_1^\pm) = 300$ GeV, BR($t\tilde{\chi}_1^\pm$) = 1	1708.09266 ATLAS-CONF-2019-015
	$\tilde{b}_1\tilde{b}_1, \tilde{b}_1 \rightarrow b\tilde{\chi}_2^0 \rightarrow bh\tilde{\chi}_1^0$	0 e, μ	6 b	E_T^{miss}	139	\tilde{b}_1	0.23-0.48	$\Delta m(\tilde{\chi}_2^0, \tilde{\chi}_1^0) = 130$ GeV, $m(\tilde{\chi}_1^0) = 100$ GeV $\Delta m(\tilde{\chi}_2^0, \tilde{\chi}_1^0) = 130$ GeV, $m(\tilde{\chi}_1^0) = 0$ GeV	SUSY-2018-31 SUSY-2018-31
	$\tilde{t}_1\tilde{t}_1, \tilde{t}_1 \rightarrow Wb\tilde{\chi}_1^0$ or $t\tilde{\chi}_1^0$	0-2 e, μ	0-2 jets/1-2 b	E_T^{miss}	36.1	\tilde{t}_1	1.0	$m(\tilde{\chi}_1^0) = 1$ GeV	1506.08616, 1709.04183, 1711.11520
	$\tilde{t}_1\tilde{t}_1, \tilde{t}_1 \rightarrow Wb\tilde{\chi}_1^0$	1 e, μ	3 jets/1 b	E_T^{miss}	139	\tilde{t}_1	0.44-0.59	$m(\tilde{\chi}_1^0) = 400$ GeV	ATLAS-CONF-2019-017
	$\tilde{t}_1\tilde{t}_1, \tilde{t}_1 \rightarrow \tau b\nu, \tilde{t}_1 \rightarrow \tau\tilde{G}$	1 $\tau + 1 e, \mu, \tau$	2 jets/1 b	E_T^{miss}	36.1	\tilde{t}_1	1.16	$m(\tilde{\tau}_1) = 800$ GeV	1803.10178
	$\tilde{t}_1\tilde{t}_1, \tilde{t}_1 \rightarrow c\tilde{\chi}_1^0 / \tilde{c}\tilde{c}, \tilde{c} \rightarrow c\tilde{\chi}_1^0$	0 e, μ	2 c	E_T^{miss}	36.1	\tilde{t}_1	0.46 0.85	$m(\tilde{\chi}_1^0) = 0$ GeV	1805.01649
	$\tilde{t}_1\tilde{t}_1, \tilde{t}_1 \rightarrow c\tilde{\chi}_1^0 / \tilde{c}\tilde{c}, \tilde{c} \rightarrow c\tilde{\chi}_1^0$	0 e, μ	mono-jet	E_T^{miss}	36.1	\tilde{t}_1	0.43	$m(\tilde{t}_1, \tilde{c}) - m(\tilde{\chi}_1^0) = 50$ GeV $m(\tilde{t}_1, \tilde{c}) - m(\tilde{\chi}_1^0) = 5$ GeV	1805.01649 1711.03301
$\tilde{t}_2\tilde{t}_2, \tilde{t}_2 \rightarrow \tilde{t}_1 + h$	1-2 e, μ	4 b	E_T^{miss}	36.1	\tilde{t}_2	0.32-0.88	$m(\tilde{\chi}_1^0) = 0$ GeV, $m(\tilde{t}_1) - m(\tilde{\chi}_1^0) = 180$ GeV	1706.03986	
$\tilde{t}_2\tilde{t}_2, \tilde{t}_2 \rightarrow \tilde{t}_1 + Z$	3 e, μ	1 b	E_T^{miss}	139	\tilde{t}_2	0.86	$m(\tilde{\chi}_1^0) = 360$ GeV, $m(\tilde{t}_1) - m(\tilde{\chi}_1^0) = 40$ GeV	ATLAS-CONF-2019-016	
EW direct	$\tilde{\chi}_1^+\tilde{\chi}_2^0$ via WZ	2-3 e, μ $ee, \mu\mu$	≥ 1	E_T^{miss} E_T^{miss}	36.1 139	$\tilde{\chi}_1^+\tilde{\chi}_2^0$ $\tilde{\chi}_1^+\tilde{\chi}_2^0$	0.6 0.205	$m(\tilde{\chi}_1^0) = 0$ $m(\tilde{\chi}_1^+) - m(\tilde{\chi}_1^0) = 5$ GeV	1403.5294, 1806.02293 ATLAS-CONF-2019-014
	$\tilde{\chi}_1^+\tilde{\chi}_1^0$ via WW	2 e, μ		E_T^{miss}	139	$\tilde{\chi}_1^+$	0.42	$m(\tilde{\chi}_1^0) = 0$	ATLAS-CONF-2019-008
	$\tilde{\chi}_1^+\tilde{\chi}_2^0$ via Wh	0-1 e, μ	2 $b/2 \gamma$	E_T^{miss}	139	$\tilde{\chi}_1^+\tilde{\chi}_2^0$	0.74	$m(\tilde{\chi}_1^0) = 70$ GeV	ATLAS-CONF-2019-019, ATLAS-CONF-2019-XYZ
	$\tilde{\chi}_1^+\tilde{\chi}_1^0$ via $\tilde{\ell}_L/\tilde{\nu}$	2 e, μ		E_T^{miss}	139	$\tilde{\chi}_1^+$	1.0	$m(\tilde{\ell}, \tilde{\nu}) = 0.5(m(\tilde{\chi}_1^+) + m(\tilde{\chi}_1^0))$	ATLAS-CONF-2019-008
	$\tilde{\tau}\tilde{\tau}, \tilde{\tau} \rightarrow \tau\tilde{\chi}_1^0$	2 τ		E_T^{miss}	139	$\tilde{\tau}$	0.16-0.3 0.12-0.39	$m(\tilde{\chi}_1^0) = 0$	ATLAS-CONF-2019-018
	$\tilde{\ell}_{1,R}\tilde{\ell}_{1,R}, \tilde{\ell} \rightarrow \ell\tilde{\chi}_1^0$	2 e, μ 2 e, μ	0 jets ≥ 1	E_T^{miss} E_T^{miss}	139 139	$\tilde{\ell}$	0.7	$m(\tilde{\chi}_1^0) = 0$	ATLAS-CONF-2019-008 ATLAS-CONF-2019-014
$\tilde{H}\tilde{H}, \tilde{H} \rightarrow h\tilde{G}/Z\tilde{G}$	0 e, μ 4 e, μ	$\geq 3 b$ 0 jets	E_T^{miss} E_T^{miss}	36.1 36.1	\tilde{H}	0.13-0.23 0.3	BR($\tilde{\chi}_1^0 \rightarrow h\tilde{G}$) = 1 BR($\tilde{\chi}_1^0 \rightarrow Z\tilde{G}$) = 1	1806.04030 1804.03602	
Long-lived particles	Direct $\tilde{\chi}_1^+\tilde{\chi}_1^-$ prod., long-lived $\tilde{\chi}_1^\pm$	Disapp. trk	1 jet	E_T^{miss}	36.1	$\tilde{\chi}_1^\pm$	0.46	Pure Wino Pure Higgsino	1712.02118 ATL-PHYS-PUB-2017-019
	Stable \tilde{g} R-hadron	Multiple	Multiple		36.1	\tilde{g}	2.0		1902.01636, 1808.04095
	Metastable \tilde{g} R-hadron, $\tilde{g} \rightarrow qq\tilde{\chi}_1^0$	Multiple	Multiple		36.1	\tilde{g} [$\tau(\tilde{g}) = 10$ ns, 0.2 ns]	2.05 2.4	$m(\tilde{\chi}_1^0) = 100$ GeV	1710.04901, 1808.04095
RPV	LFV $pp \rightarrow \tilde{\nu}_\tau + X, \tilde{\nu}_\tau \rightarrow e\mu/\ell\tau/\mu\tau$	$e\mu, e\tau, \mu\tau$			3.2	$\tilde{\nu}_\tau$	1.9	$\lambda'_{311} = 0.11, \lambda'_{132/133/233} = 0.07$	1607.08079
	$\tilde{\chi}_1^+\tilde{\chi}_1^0/\tilde{\chi}_2^0 \rightarrow WW/Z\ell\ell\nu\nu$	4 e, μ	0 jets	E_T^{miss}	36.1	$\tilde{\chi}_1^+\tilde{\chi}_2^0$ [$\lambda'_{333} \neq 0, \lambda'_{124} \neq 0$]	0.82 1.33	$m(\tilde{\chi}_1^0) = 100$ GeV	1804.03602
	$\tilde{g}\tilde{g}, \tilde{g} \rightarrow qq\tilde{\chi}_1^0, \tilde{\chi}_1^0 \rightarrow qq\tilde{\chi}_1^0$	4-5 large- R jets Multiple	Multiple		36.1 36.1	$\tilde{g}\tilde{g}$ [$m(\tilde{\chi}_1^0) = 200$ GeV, 1100 GeV] $\tilde{g}\tilde{g}$ [$\lambda'_{112} = 2e-4, 2e-5$]	1.3 1.05 2.0	Large λ'_{112} $m(\tilde{\chi}_1^0) = 200$ GeV, bino-like	1804.03568 ATLAS-CONF-2018-003
	$\tilde{u}, \tilde{t} \rightarrow \tilde{\chi}_1^0, \tilde{\chi}_1^0 \rightarrow tbs$	Multiple	Multiple		36.1	\tilde{g} [$\lambda'_{323} = 2e-4, 1e-2$]	0.55 1.05	$m(\tilde{\chi}_1^0) = 200$ GeV, bino-like	ATLAS-CONF-2018-003
	$\tilde{t}_1\tilde{t}_1, \tilde{t}_1 \rightarrow bs$	2 jets + 2 b			36.7	\tilde{t}_1 [qq, bs]	0.42 0.61		1710.07171
$\tilde{t}_1\tilde{t}_1, \tilde{t}_1 \rightarrow q\ell$	2 e, μ 1 μ	2 b DV		36.1 136	\tilde{t}_1	1.0	BR($\tilde{t}_1 \rightarrow b\ell/\mu\mu$) > 20% BR($\tilde{t}_1 \rightarrow q\mu$) = 100%, $\cos\theta = 1$	1710.05544 ATLAS-CONF-2019-006	

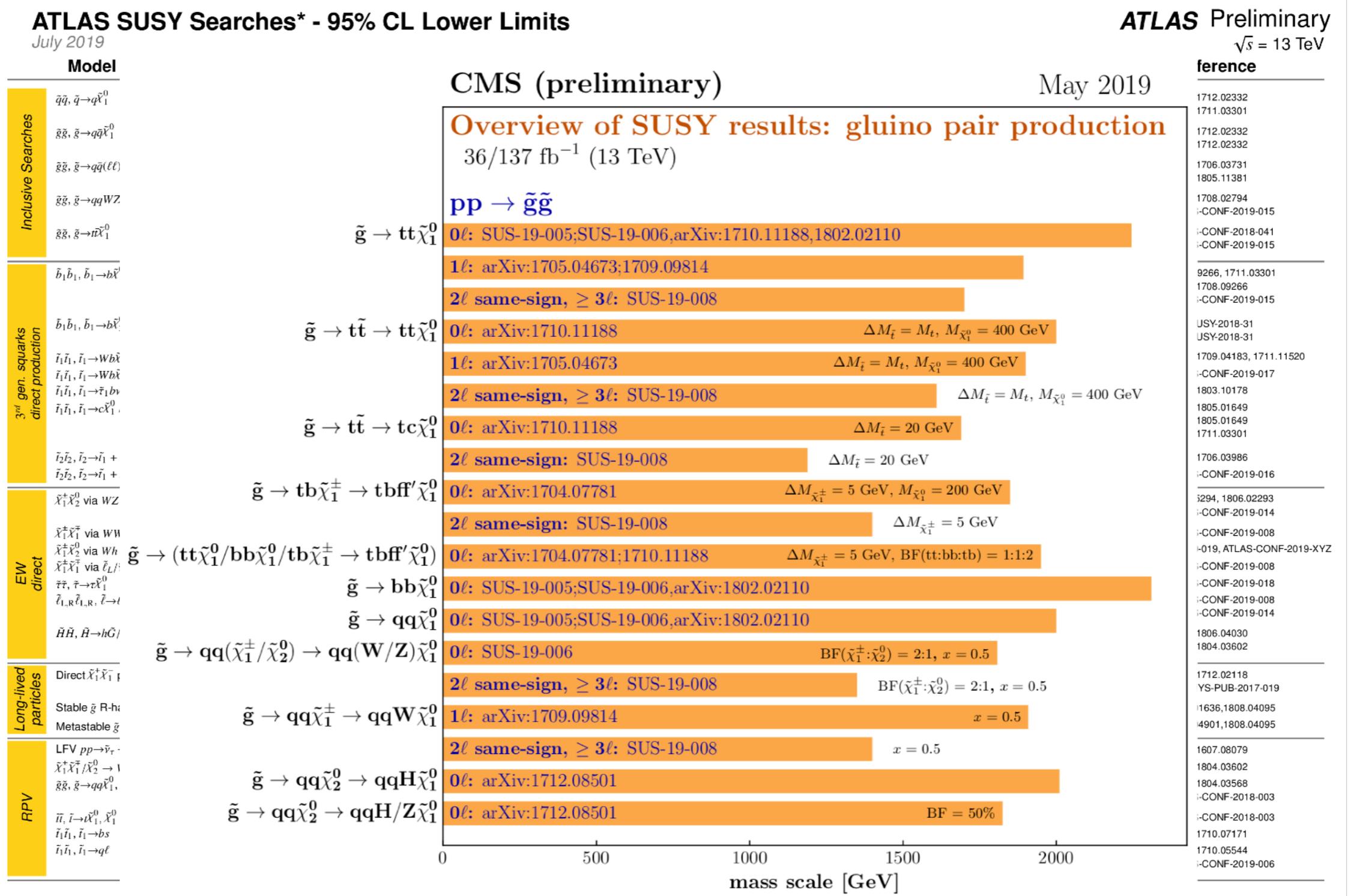
*Only a selection of the available mass limits on new states or phenomena is shown. Many of the limits are based on

10⁻¹

1

Mass scale [TeV]

Current Status of NP Searches @ LHC



*Only a selection of phenomena is shown. Selection of observed limits at 95% C.L. (theory uncertainties are not included). Probe up to the quoted mass limit for light LSPs unless stated otherwise. The quantities ΔM and x represent the absolute mass difference between the primary sparticle and the LSP, and the difference between the intermediate

Current Status of NP Searches @ LHC

ATLAS SUSY Searches* - 95% CL Lower Limits

July 2019

Model

$\tilde{q}\tilde{q}, \tilde{q} \rightarrow q\tilde{\chi}_1^0$

$\tilde{g}\tilde{g}, \tilde{g} \rightarrow q\tilde{q}\tilde{\chi}_1^0$

Searches

CMS (preliminary)

Overview of SUSY results: gluino pair production

May 2019

ATLAS Preliminary

$\sqrt{s} = 13$ TeV

Reference

1712.02332

1711.03301

1712.02332

1712.02332

ATLAS Exotics Searches* - 95% CL Upper Exclusion Limits

Status: May 2019

ATLAS Preliminary

$\int \mathcal{L} dt = (3.2 - 139) \text{ fb}^{-1}$

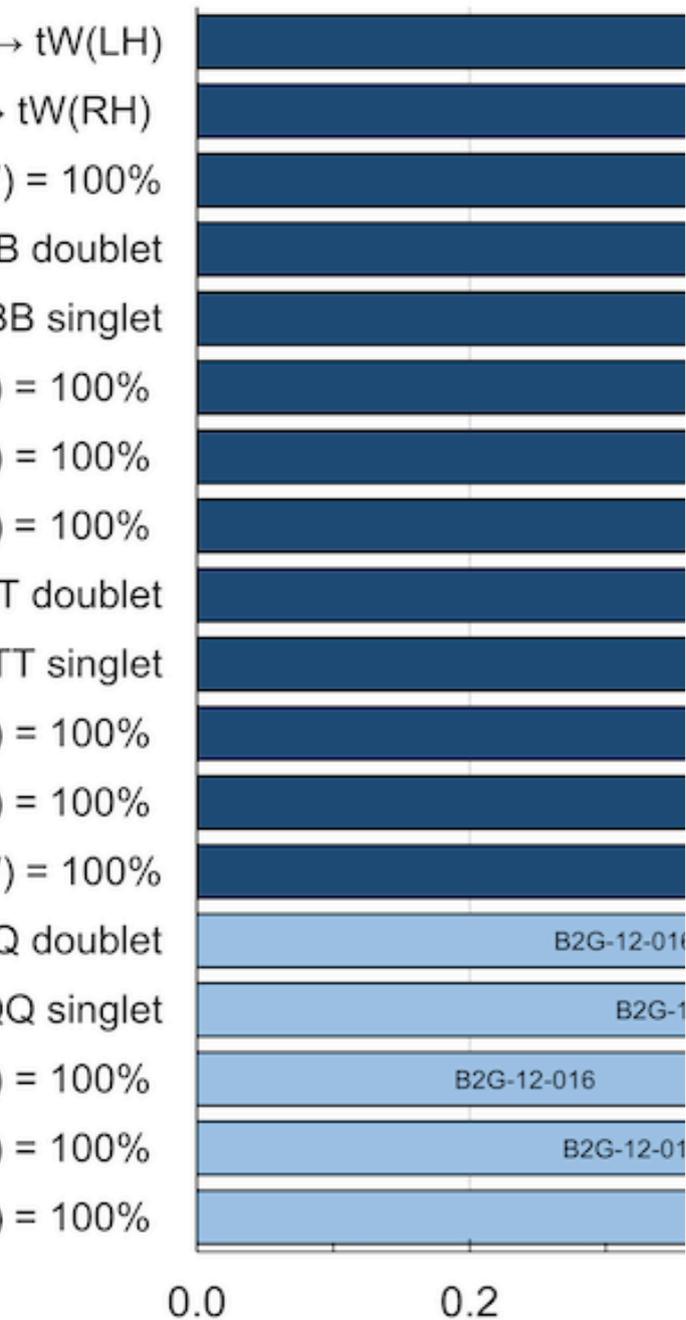
$\sqrt{s} = 8, 13$ TeV

Model	ℓ, γ	Jets [†]	E_T^{miss}	$\int \mathcal{L} dt [\text{fb}^{-1}]$	Limit	Reference		
Extra dimensions	ADD $G_{KK} + g/q$	0 e, μ	1-4 j	Yes	36.1	M_D 7.7 TeV	$n = 2$	1711.03301
	ADD non-resonant $\gamma\gamma$	2 γ	-	-	36.7	M_S 8.6 TeV	$n = 3$ HLZ NLO	1707.04147
	ADD QBH	-	2 j	-	37.0	M_{th} 8.9 TeV	$n = 6$	1703.09127
	ADD BH high $\sum p_T$	$\geq 1 e, \mu$	$\geq 2 j$	-	3.2	M_{th} 8.2 TeV	$n = 6, M_D = 3 \text{ TeV, rot BH}$	1606.02265
	ADD BH multijet	-	$\geq 3 j$	-	3.6	M_{th} 9.55 TeV	$n = 6, M_D = 3 \text{ TeV, rot BH}$	1512.02586
	RS1 $G_{KK} \rightarrow \gamma\gamma$	2 γ	-	-	36.7	G_{KK} mass 4.1 TeV	$k/\bar{M}_{Pl} = 0.1$	1707.04147
	Bulk RS $G_{KK} \rightarrow WW/ZZ$	multi-channel	-	-	36.1	G_{KK} mass 2.3 TeV	$k/\bar{M}_{Pl} = 1.0$	1808.02380
	Bulk RS $G_{KK} \rightarrow WW \rightarrow qq\bar{q}\bar{q}$	0 e, μ	2 J	-	139	G_{KK} mass 1.6 TeV	$k/\bar{M}_{Pl} = 1.0$	ATLAS-CONF-2019-003
	Bulk RS $G_{KK} \rightarrow tt$	1 e, μ	$\geq 1 b, \geq 1J/2j$	Yes	36.1	g_{KK} mass 3.8 TeV	$\Gamma/m = 15\%$	1804.10823
	2UED / RPP	1 e, μ	$\geq 2 b, \geq 3 j$	Yes	36.1	KK mass 1.8 TeV	Tier (1,1), $\mathcal{B}(A^{(1,1)} \rightarrow tt) = 1$	1803.09678
Gauge bosons	SSM $Z' \rightarrow \ell\ell$	2 e, μ	-	-	139	Z' mass 5.1 TeV		1903.06248
	SSM $Z' \rightarrow \tau\tau$	2 τ	-	-	36.1	Z' mass 2.42 TeV		1709.07242
	Leptophobic $Z' \rightarrow bb$	-	2 b	-	36.1	Z' mass 2.1 TeV		1805.09299
	Leptophobic $Z' \rightarrow tt$	1 e, μ	$\geq 1 b, \geq 1J/2j$	Yes	36.1	Z' mass 3.0 TeV	$\Gamma/m = 1\%$	1804.10823
	SSM $W' \rightarrow \ell\nu$	1 e, μ	-	Yes	139	W' mass 6.0 TeV		CERN-EP-2019-100
	SSM $W' \rightarrow \tau\nu$	1 τ	-	Yes	36.1	W' mass 3.7 TeV		1801.06992
	HVT $V' \rightarrow WZ \rightarrow qq\bar{q}\bar{q}$ model B	0 e, μ	2 J	-	139	V' mass 3.6 TeV	$g_V = 3$	ATLAS-CONF-2019-003
	HVT $V' \rightarrow WH/ZH$ model B	multi-channel	-	-	36.1	V' mass 2.93 TeV	$g_V = 3$	1712.06518
	LRSM $W_R \rightarrow tb$	multi-channel	-	-	36.1	W_R mass 3.25 TeV		1807.10473
	LRSM $W_R \rightarrow \mu N_R$	2 μ	1 J	-	80	W_R mass 5.0 TeV	$m(N_R) = 0.5 \text{ TeV, } g_L = g_R$	1904.12679
CI	CI $qq\bar{q}\bar{q}$	-	2 j	-	37.0	Λ 21.8 TeV	η_{LL}^-	1703.09127
	CI $\ell\ell q\bar{q}$	2 e, μ	-	-	36.1	Λ 40.0 TeV	η_{LL}^-	1707.02424
	CI $tt\bar{t}\bar{t}$	$\geq 1 e, \mu$	$\geq 1 b, \geq 1 j$	Yes	36.1	Λ 2.57 TeV	$ C_{4t} = 4\pi$	1811.02305
DM	Axial-vector mediator (Dirac DM)	0 e, μ	1-4 j	Yes	36.1	m_{med} 1.55 TeV	$g_q = 0.25, g_\nu = 1.0, m(\chi) = 1 \text{ GeV}$	1711.03301
	Colored scalar mediator (Dirac DM)	0 e, μ	1-4 j	Yes	36.1	m_{med} 1.67 TeV	$g = 1.0, m(\chi) = 1 \text{ GeV}$	1711.03301
	$VV\chi\chi$ EFT (Dirac DM)	0 e, μ	1 J, $\leq 1 j$	Yes	3.2	M_s 700 GeV	$m(\chi) < 150 \text{ GeV}$	1608.02372
	Scalar reson. $\phi \rightarrow t\chi$ (Dirac DM)	0-1 e, μ	1 b, 0-1 J	Yes	36.1	m_ϕ 3.4 TeV	$y = 0.4, \lambda = 0.2, m(\chi) = 10 \text{ GeV}$	1812.09743
LQ	Scalar LQ 1 st gen	1,2 e	$\geq 2 j$	Yes	36.1	LQ mass 1.4 TeV	$\beta = 1$	1902.00377
	Scalar LQ 2 nd gen	1,2 μ	$\geq 2 j$	Yes	36.1	LQ mass 1.56 TeV	$\beta = 1$	1902.00377
	Scalar LQ 3 rd gen	2 τ	2 b	-	36.1	LQ ₃ mass 1.03 TeV	$\mathcal{B}(LQ_3^+ \rightarrow b\tau) = 1$	1902.08103
	Scalar LQ 3 rd gen	0-1 e, μ	2 b	Yes	36.1	LQ ₃ mass 970 GeV	$\mathcal{B}(LQ_3^+ \rightarrow t\tau) = 0$	1902.08103
Heavy quarks	VLQ $TT \rightarrow Ht/Zt/Wb + X$	multi-channel	-	-	36.1	T mass 1.37 TeV	SU(2) doublet	1808.02343
	VLQ $BB \rightarrow Wt/Zb + X$	multi-channel	-	-	36.1	B mass 1.34 TeV	SU(2) doublet	1808.02343
	VLQ $T_{5/3} T_{5/3} T_{5/3} \rightarrow Wt + X$	2(SS) $\geq 3 e, \mu \geq 1 b, \geq 1 j$	Yes	36.1	$T_{5/3}$ mass 1.64 TeV	$\mathcal{B}(T_{5/3} \rightarrow Wt) = 1, c(T_{5/3} Wt) = 1$	1807.11883	
	VLQ $Y \rightarrow Wb + X$	1 e, μ	$\geq 1 b, \geq 1 j$	Yes	36.1	Y mass 1.85 TeV	$\mathcal{B}(Y \rightarrow Wb) = 1, c_R(Wb) = 1$	1812.07343
	VLQ $B \rightarrow Hb + X$	0 $e, \mu, 2 \gamma$	$\geq 1 b, \geq 1 j$	Yes	79.8	B mass 1.21 TeV	$\kappa_B = 0.5$	ATLAS-CONF-2018-024
	VLQ $QQ \rightarrow WqWq$	1 e, μ	$\geq 4 j$	Yes	20.3	Q mass 690 GeV		1509.04261
Excited fermions	Excited quark $q^* \rightarrow qg$	-	2 j	-	139	q^* mass 6.7 TeV	only u' and d' , $\Lambda = m(q^*)$	ATLAS-CONF-2019-007
	Excited quark $q^* \rightarrow q\gamma$	1 γ	1 j	-	36.7	q^* mass 5.3 TeV	only u' and d' , $\Lambda = m(q^*)$	1709.10440
	Excited quark $b^* \rightarrow bg$	-	1 b, 1 j	-	36.1	b^* mass 2.6 TeV		1805.09299
	Excited lepton ℓ^*	3 e, μ	-	-	20.3	ℓ^* mass 3.0 TeV	$\Lambda = 3.0 \text{ TeV}$	1411.2921
	Excited lepton ν^*	3 e, μ, τ	-	-	20.3	ν^* mass 1.6 TeV	$\Lambda = 1.6 \text{ TeV}$	1411.2921
Other	Type III Seesaw	1 e, μ	$\geq 2 j$	Yes	79.8	N^0 mass 560 GeV		ATLAS-CONF-2018-020
	LRSM Majorana ν	2 μ	2 j	-	36.1	N_R mass 3.2 TeV	$m(W_R) = 4.1 \text{ TeV, } g_L = g_R$	1809.11105
	Higgs triplet $H^{++} \rightarrow \ell\ell$	2,3,4 e, μ (SS)	-	-	36.1	H^{++} mass 870 GeV	DY production	1710.09748
	Higgs triplet $H^{++} \rightarrow \ell\tau$	3 e, μ, τ	-	-	20.3	H^{++} mass 400 GeV	DY production, $\mathcal{B}(H^{++} \rightarrow \ell\tau) = 1$	1411.2921
	Multi-charged particles	-	-	-	36.1	multi-charged particle mass 1.22 TeV	DY production, $ q = 5e$	1812.03673

Vector-like quark single production

$\sigma(Xtq) \times B(X \rightarrow tW)$, RH	B2G-17-01
$\sigma(Xtq) \times B(X \rightarrow tW)$, LH	B2G-17-01
$\sigma(Btq) \times B(B \rightarrow tW)$, RH	B2G-17-01
$\sigma(Bbq) \times B(B \rightarrow tW)$, RH	B2G-17-01

Ve

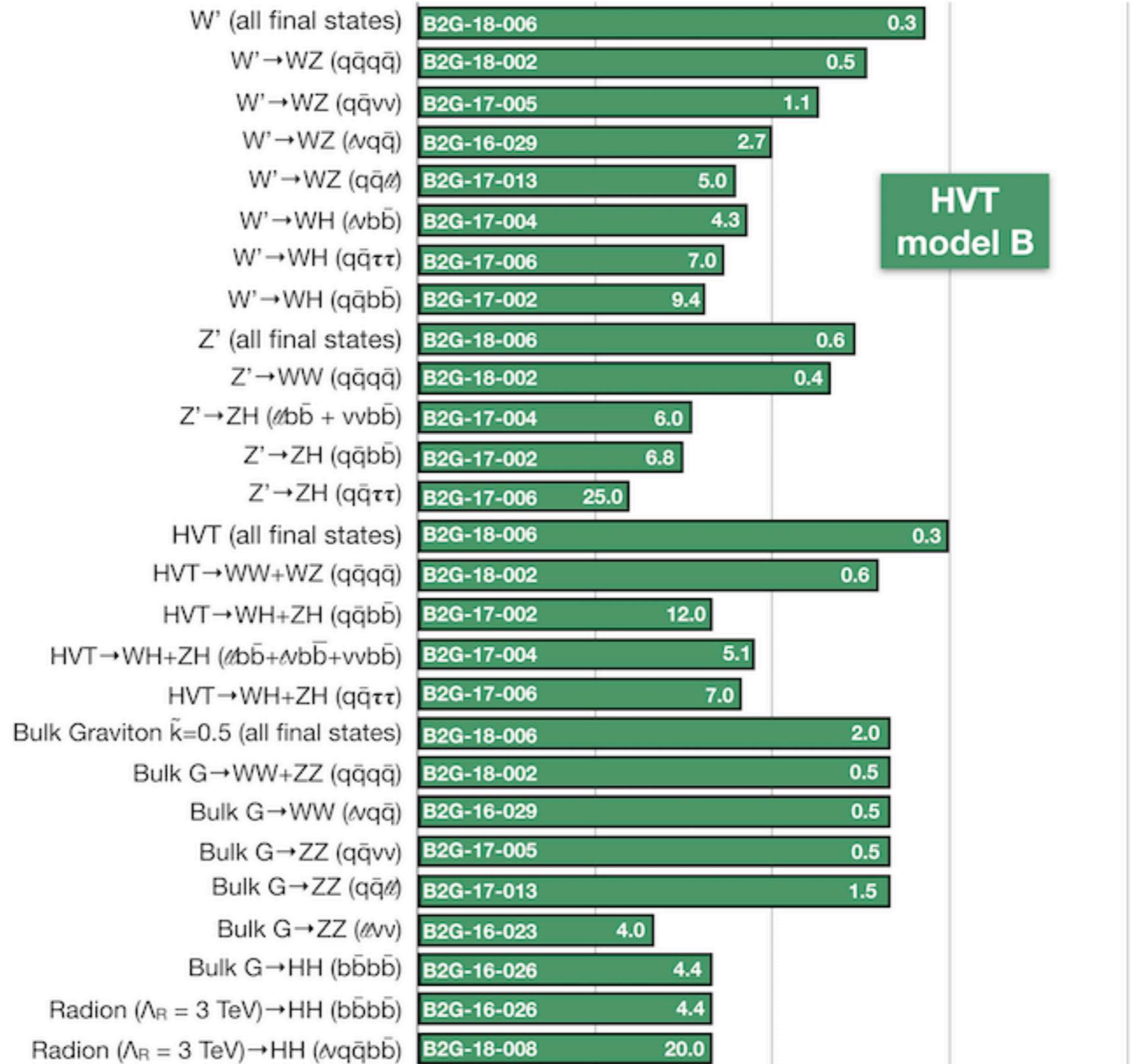


HEP 2019

Selection of observed limits at 95% C.L. The quantities ΔM and x represent the mass of the sparticle and the LSP relative to ΔM , respectively.

BB $\rightarrow (\ell^+ \ell^+)$
BB $\rightarrow (\ell^+ \ell^-)$

Resonances to dibosons ($\sqrt{s} = 13$ TeV)



HVT model B

CMS, EPS-HEP 2019

95% CL Lower Mass Limit [TeV]

(Upper Cross Section Limit [fb])

Previous model-independent approaches

“the general search”

$$R(x) = \frac{\mathcal{L}(x|data)}{\mathcal{L}(x|B_{sim})}$$

Previous model-independent approaches

“*the general search*”

$$R(x) = \frac{\mathcal{L}(x|data)}{\mathcal{L}(x|B_{sim})}$$

x: a single feature (almost always binned)

Idea: compare data vs **simulated** SM background in ID histograms.

Previous model-independent approaches

“*the general search*”

fully signal model
independent
background model
dependent

$$R(x) = \frac{\mathcal{L}(x|data)}{\mathcal{L}(x|B_{sim})}$$

x: a single feature (almost
always binned)

Idea: compare data vs **simulated** SM background in 1D histograms.

Previous model-independent approaches

“*the general search*”

fully signal model
independent
background model
dependent

$$R(x) = \frac{\mathcal{L}(x|data)}{\mathcal{L}(x|B_{sim})}$$

x: a single feature (almost always binned)

Idea: compare data vs **simulated** SM background in 1D histograms.

[Can also imagine turbocharged version: train DNN on full phase space to distinguish data from background MC (D’Agnolo, Wulzer et al 1806.02350, 1912.12155)]

Previous model-independent approaches

“*the general search*”

fully signal model
independent
background model
dependent

$$R(x) = \frac{\mathcal{L}(x|data)}{\mathcal{L}(x|B_{sim})}$$

x: a single feature (almost always binned)

Idea: compare data vs **simulated** SM background in 1D histograms.

[Can also imagine turbocharged version: train DNN on full phase space to distinguish data from background MC (D’Agnolo, Wulzer et al 1806.02350, 1912.12155)]

Data vs bg likelihood ratio: not optimal for any specific signal, but could be optimal for rejecting background hypothesis.

Previous model-independent approaches

“the general search”

fully signal model
independent
background model
dependent

$$R(x) = \frac{\mathcal{L}(x|data)}{\mathcal{L}(x|B_{sim})}$$

x: a single feature (almost always binned)

Idea: compare data vs **simulated** SM background in 1D histograms.

[Can also imagine turbocharged version: train DNN on full phase space to distinguish data from background MC (D’Agnolo, Wulzer et al 1806.02350, 1912.12155)]

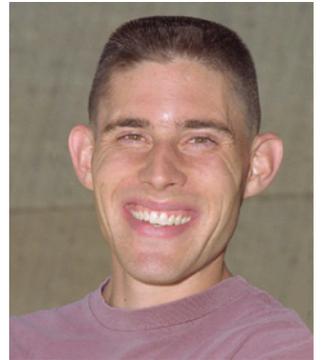
Data vs bg likelihood ratio: not optimal for any specific signal, but could be optimal for rejecting background hypothesis.

Long but somewhat neglected history of searches of this type

Previous model-independent approaches

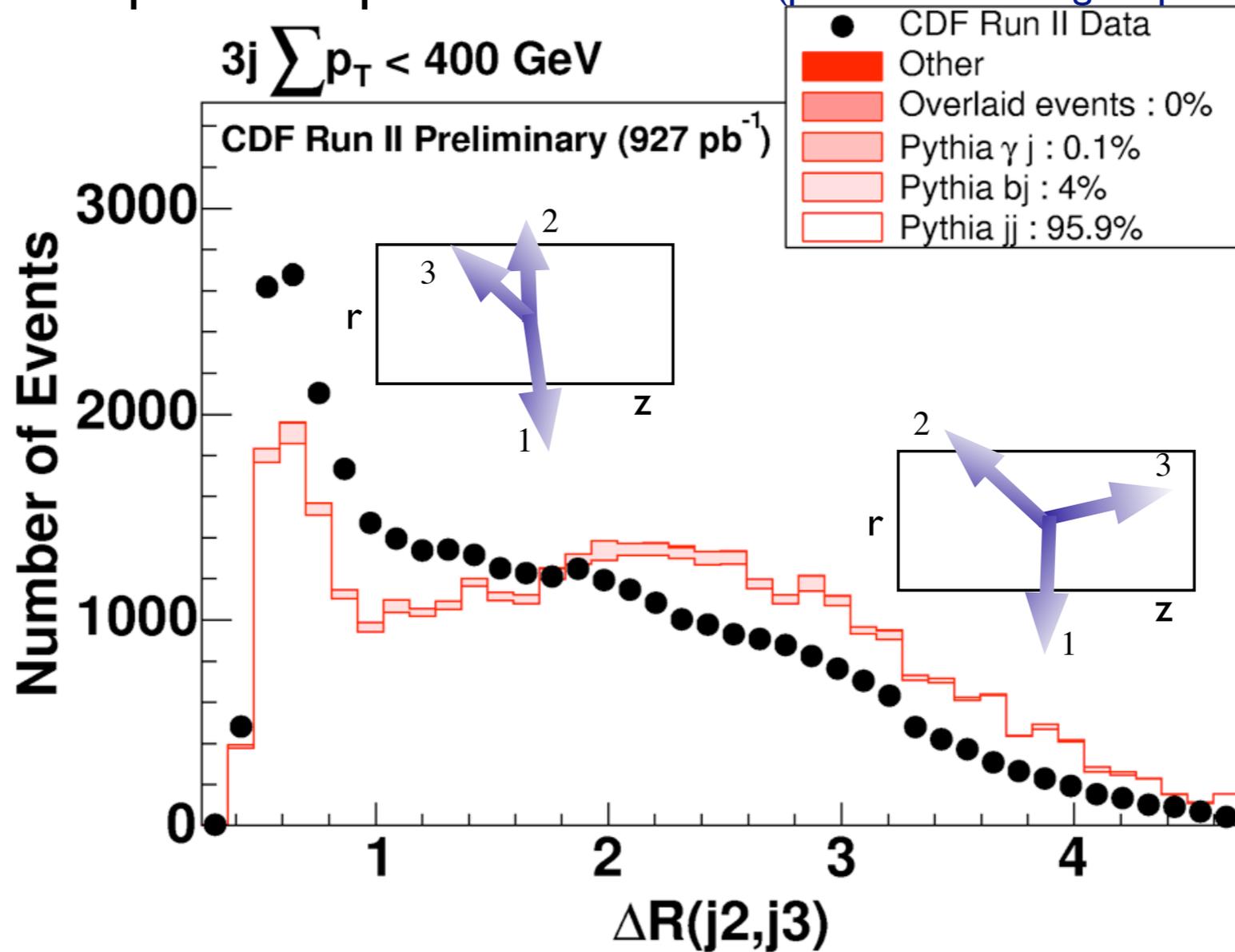
A brief history of model independent searches in HEP:

- **D0** *“Sleuth”* PRD 62:092004 (2000)
PRD 64:012004 (2001)
PRL 86:3712 (2001)
- **H1 (Hera)** *“General Search”* PLB 602:14-30 (2004)
0705.3721
- **CDF** *“Sleuth/Vista”* 0712.1311 PRD 78:012002 (2008)
0712.2534 (submitted to PRL, NEVER PUBLISHED)
0809.3781 PRD 79:011101 (2009)
- **CMS** *“MUSIC”* CMS-PAS-EXO-14-016
- **ATLAS** *“Model independent general search”* 1807.07447 EPJC 79:120 (2019)



An example of what is found

Sample discrepant distribution (parton showering suspected)



From B. Knuteson talk at UMich (2008)

Previous model-independent approaches

“the bump hunt”

partially signal and
background model
independent

$$R(m) = \frac{\mathcal{L}(m|data)}{\mathcal{L}(m|B_{data})}$$

m: a single feature

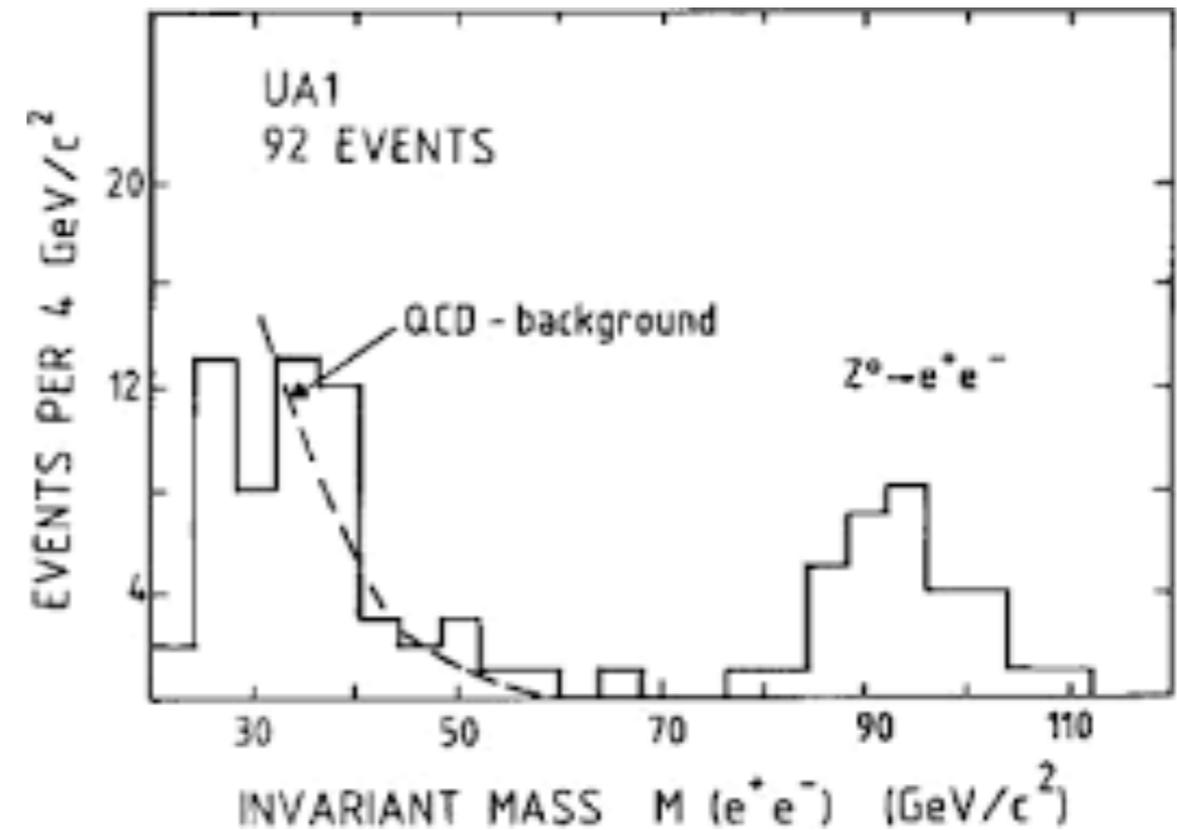
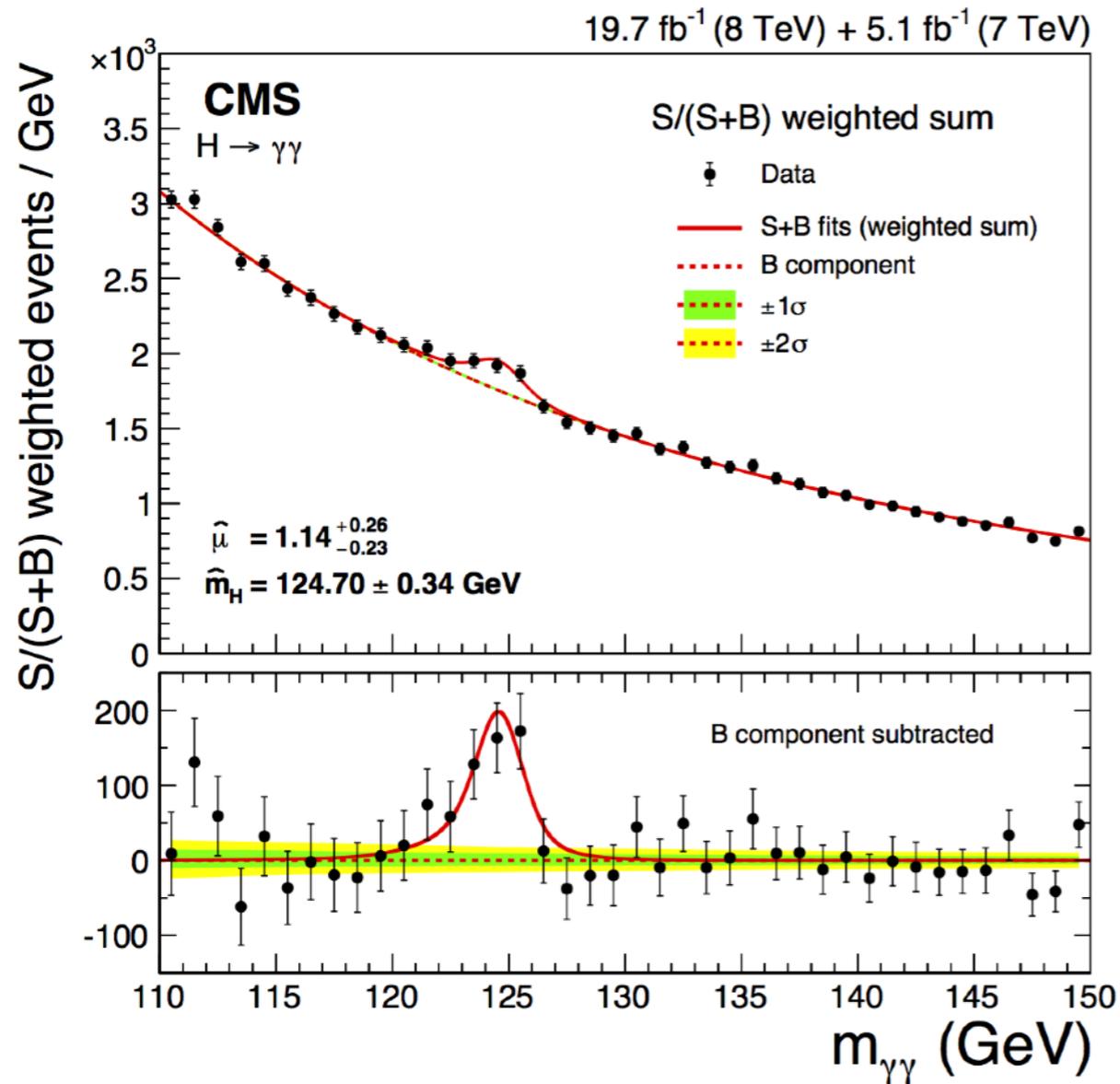
Idea: assume signal is localized in m while background is smooth.

Use **sidebands** $m \notin (m_0 - \delta m, m_0 + \delta m)$ to interpolate background into **signal region** $m \in (m_0 - \delta m, m_0 + \delta m)$.

Classic method, used in many discoveries

Previous model-independent approaches

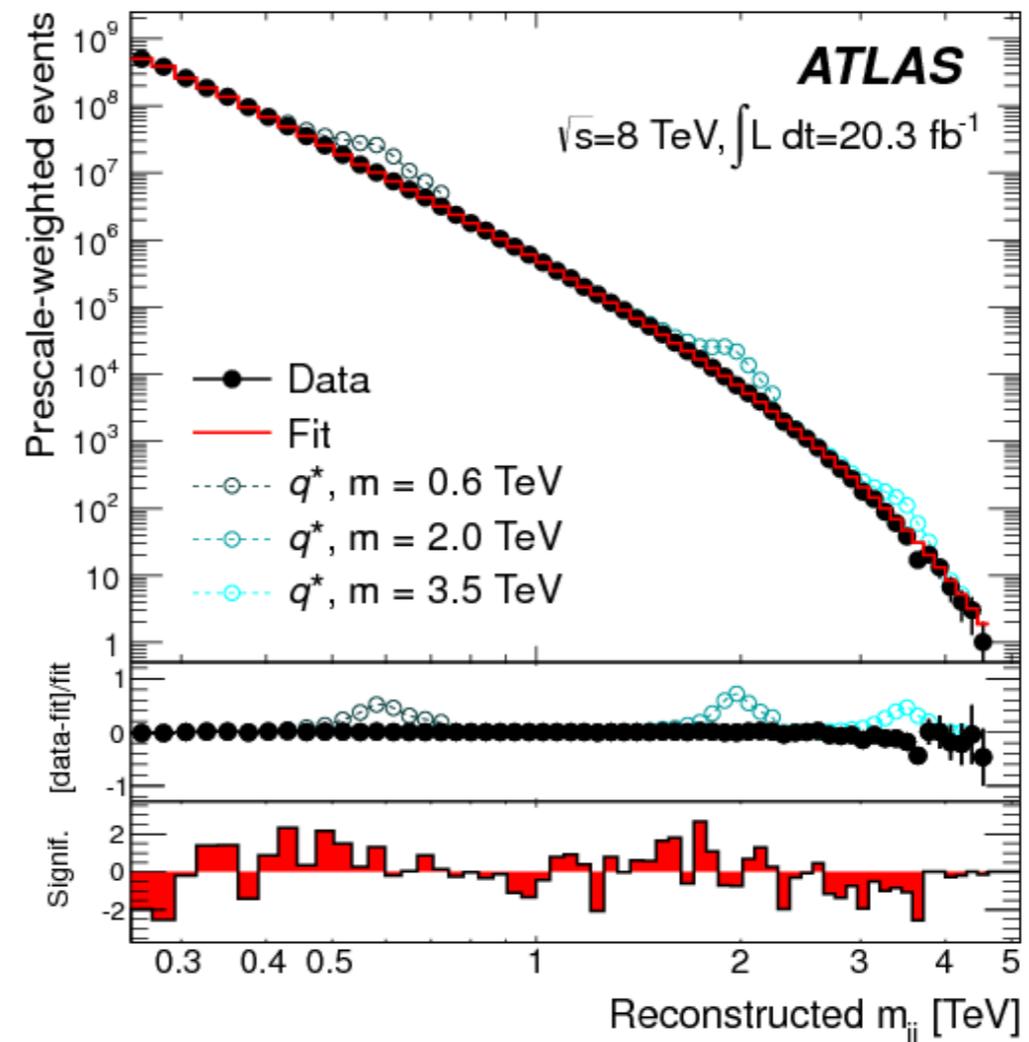
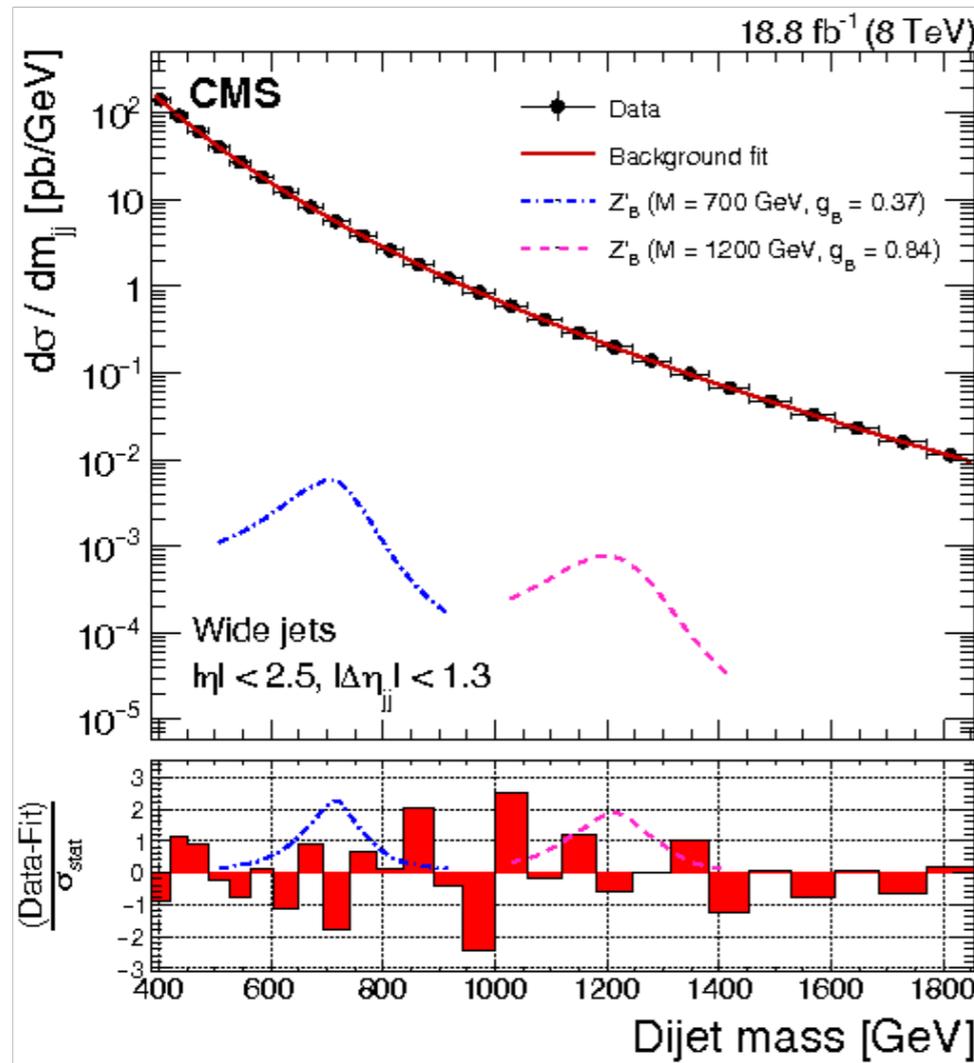
“the bump hunt”



Higgs, Z, ...

Previous model-independent approaches

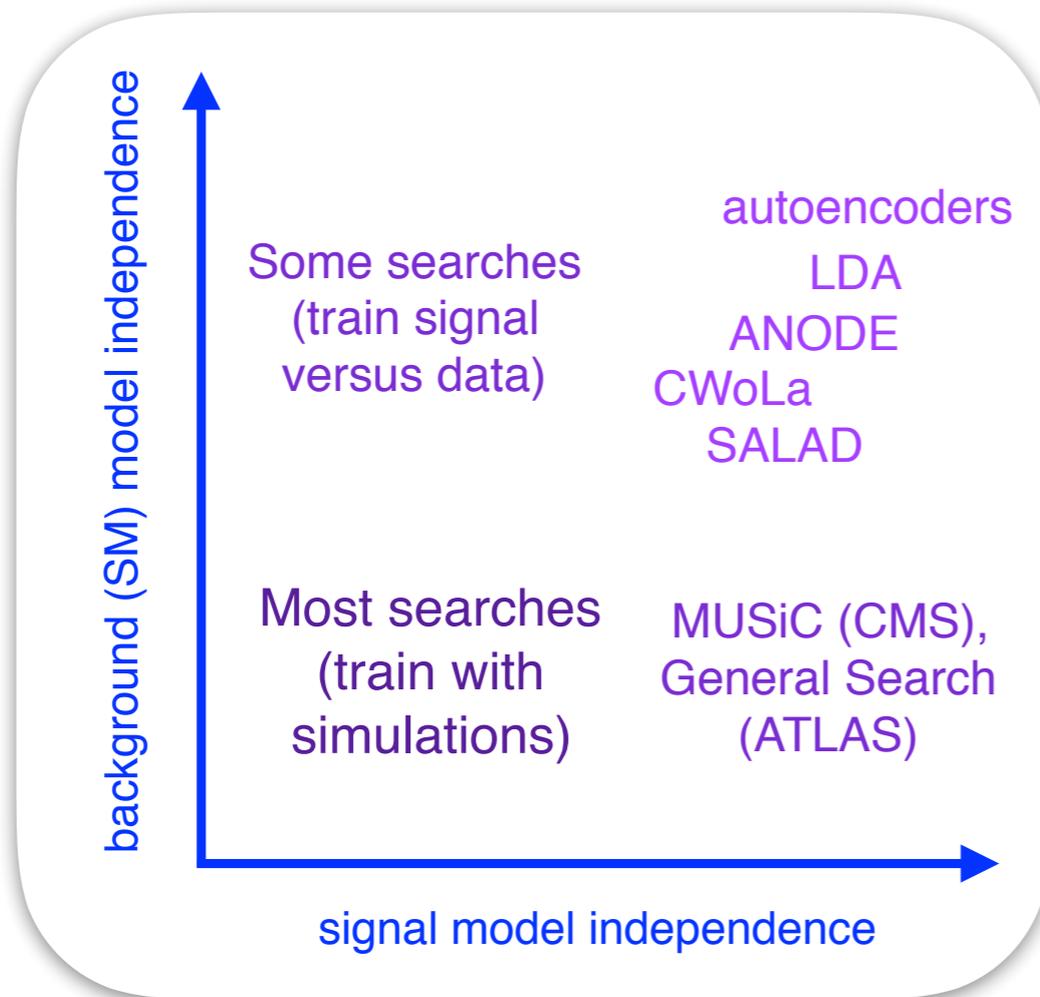
“the bump hunt”



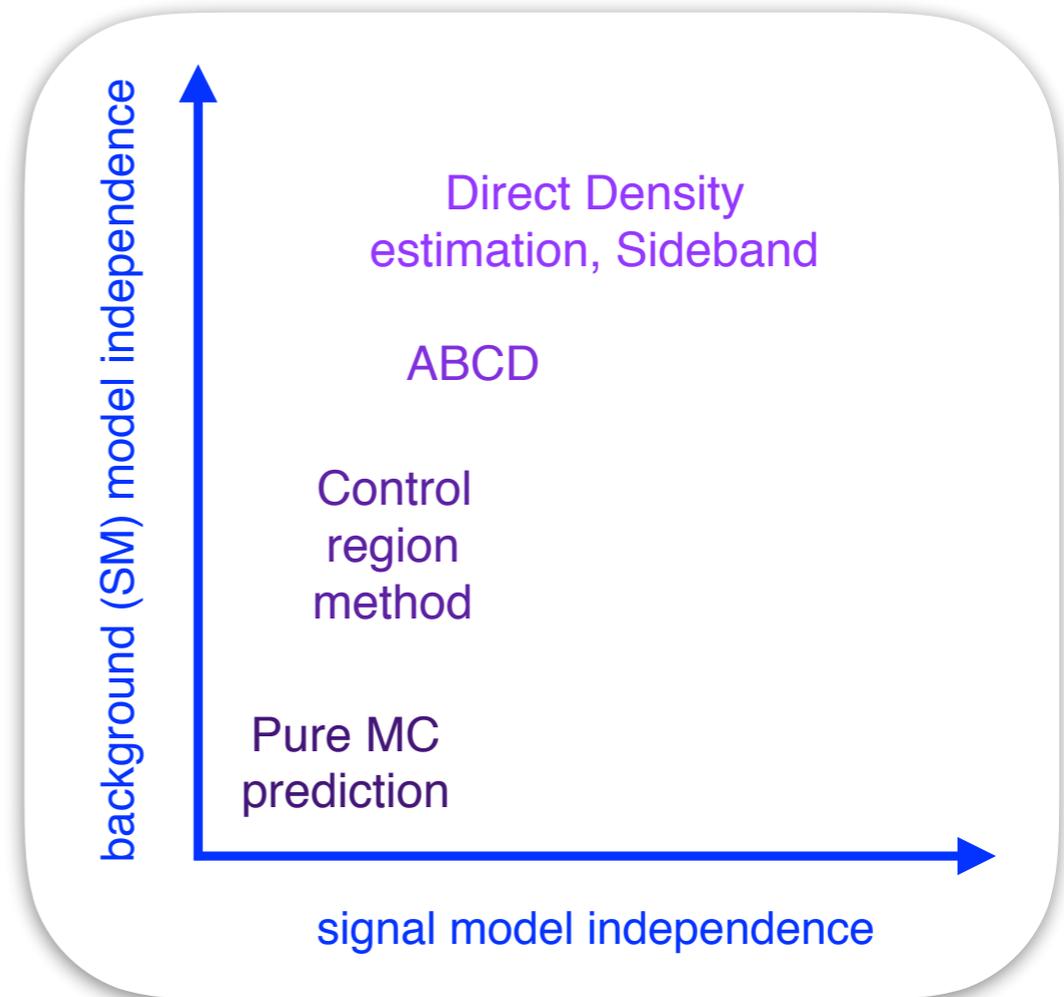
Also a classic search for new physics — e.g. a hypothetical heavy BSM particle that decays to pairs of jets

Overview of search strategies

from Nachman & DS 2001.04990

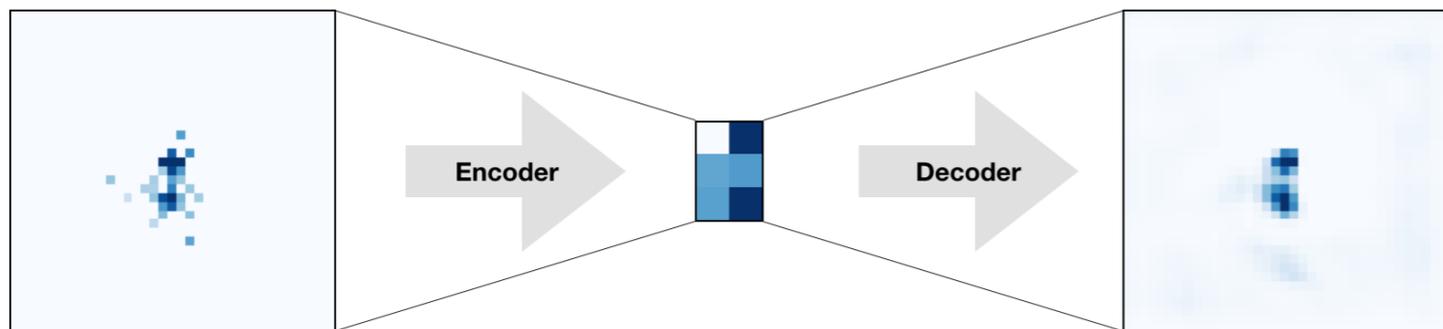


(a) Signal sensitivity

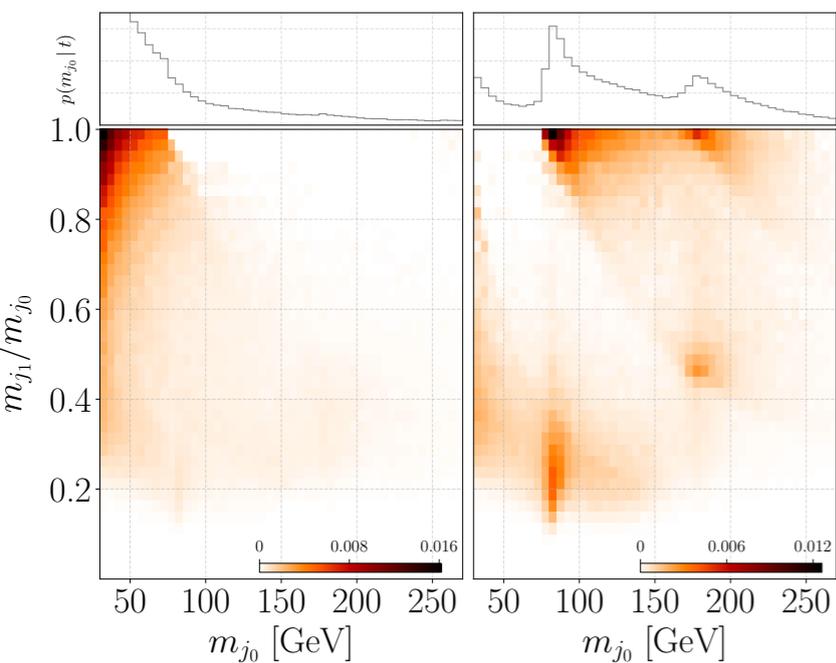


(b) Background specificity

Autoencoders

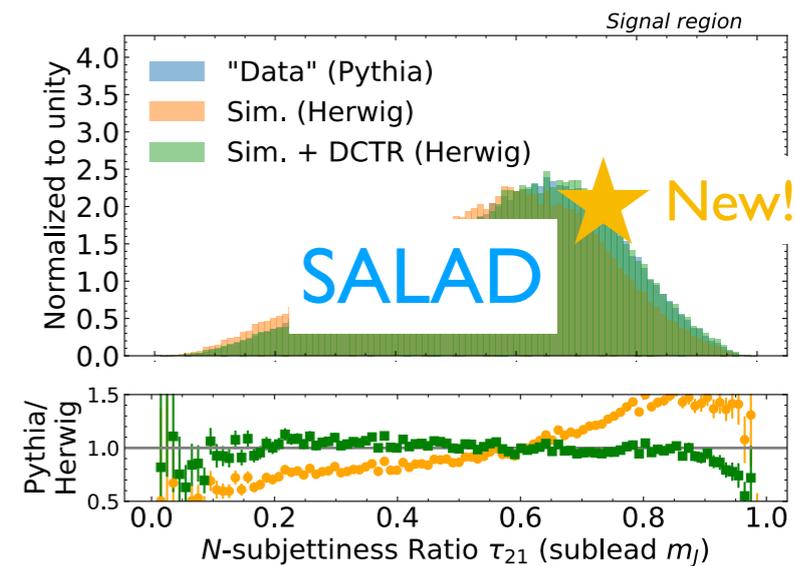
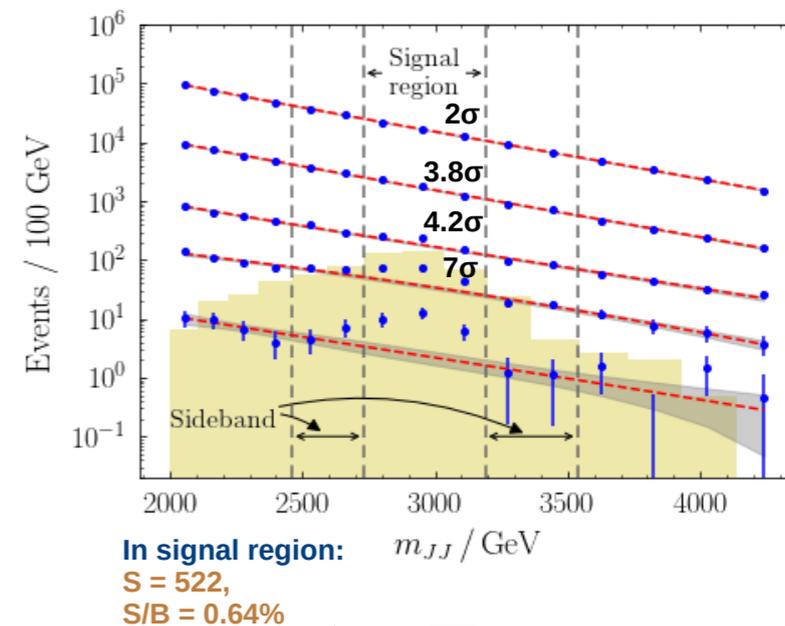


Probabilistic Modeling

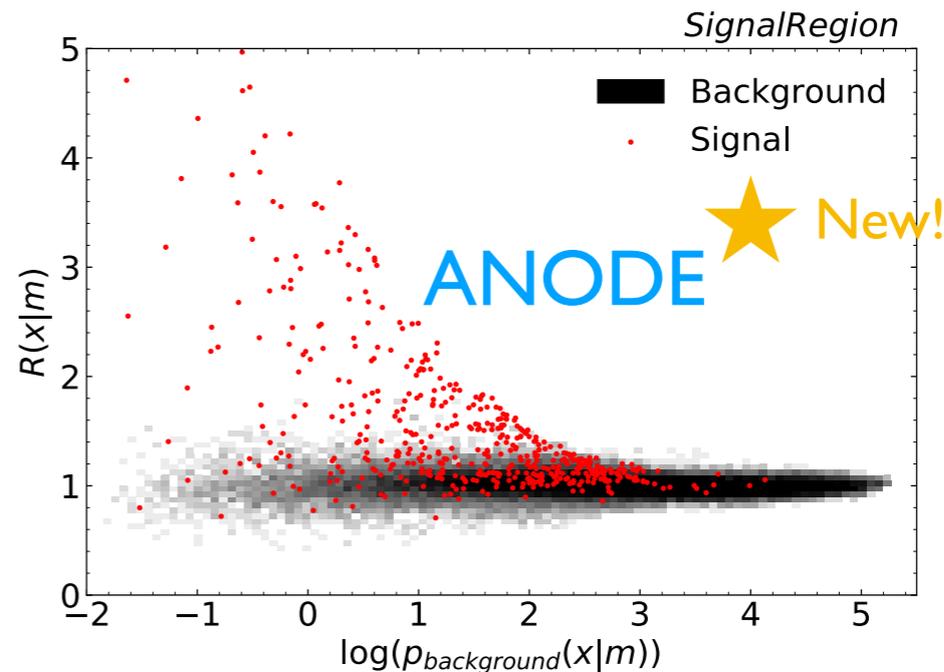
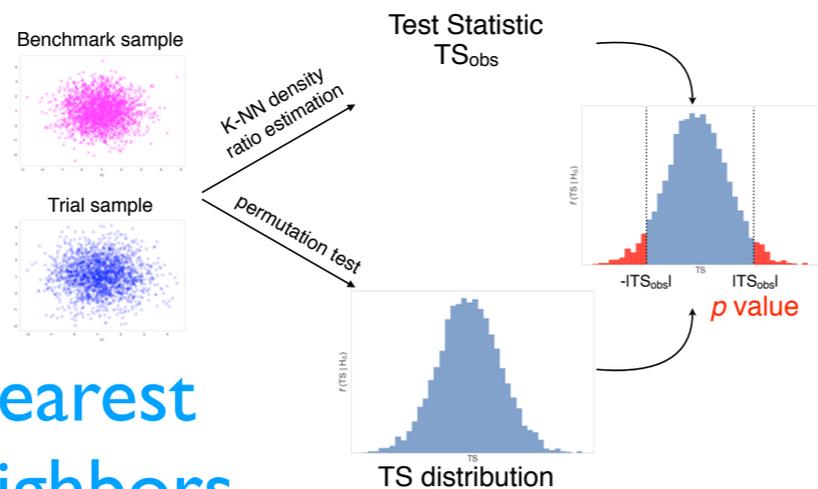


3. New approaches

CWoLa

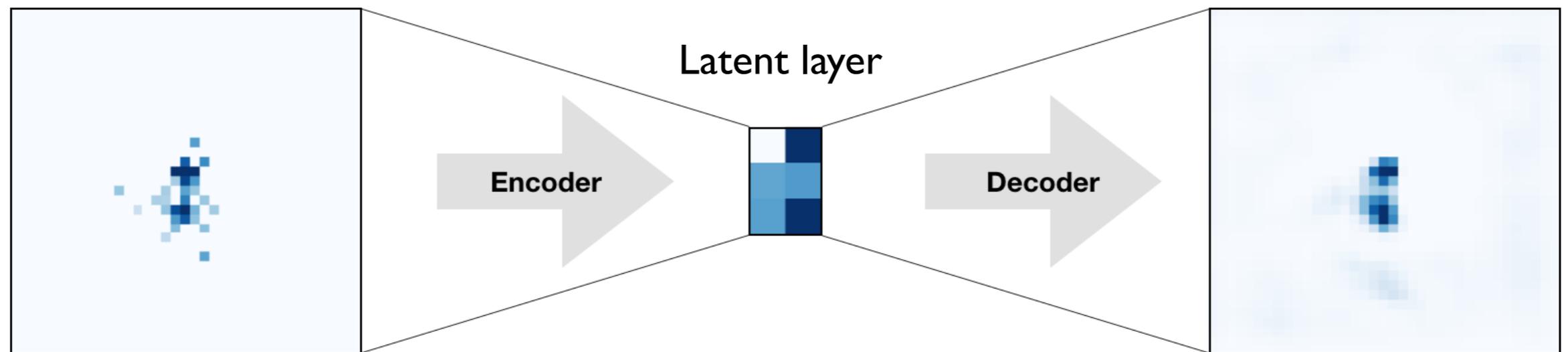


Nearest Neighbors



Searching for NP with deep autoencoders

Heimel et al I808.08979; Farina, Nakai & DS I808.08992



An autoencoder maps an input into a reduced “latent representation” and then attempts to reconstruct the original input from it.

Can use reconstruction error as an anomaly threshold!

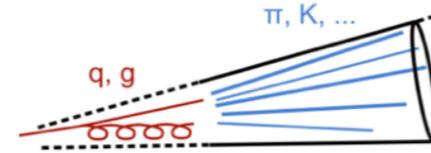
See also:

Hajer et al “Novelty Detection Meets Collider Physics” I807.10261

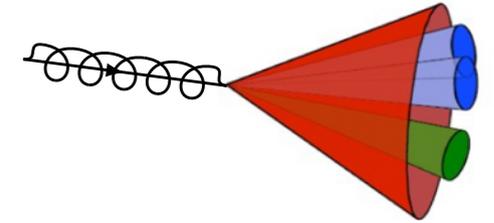
Cerri et al “Variational Autoencoders for New Physics Mining at the Large Hadron Collider” I811.10276

Performance

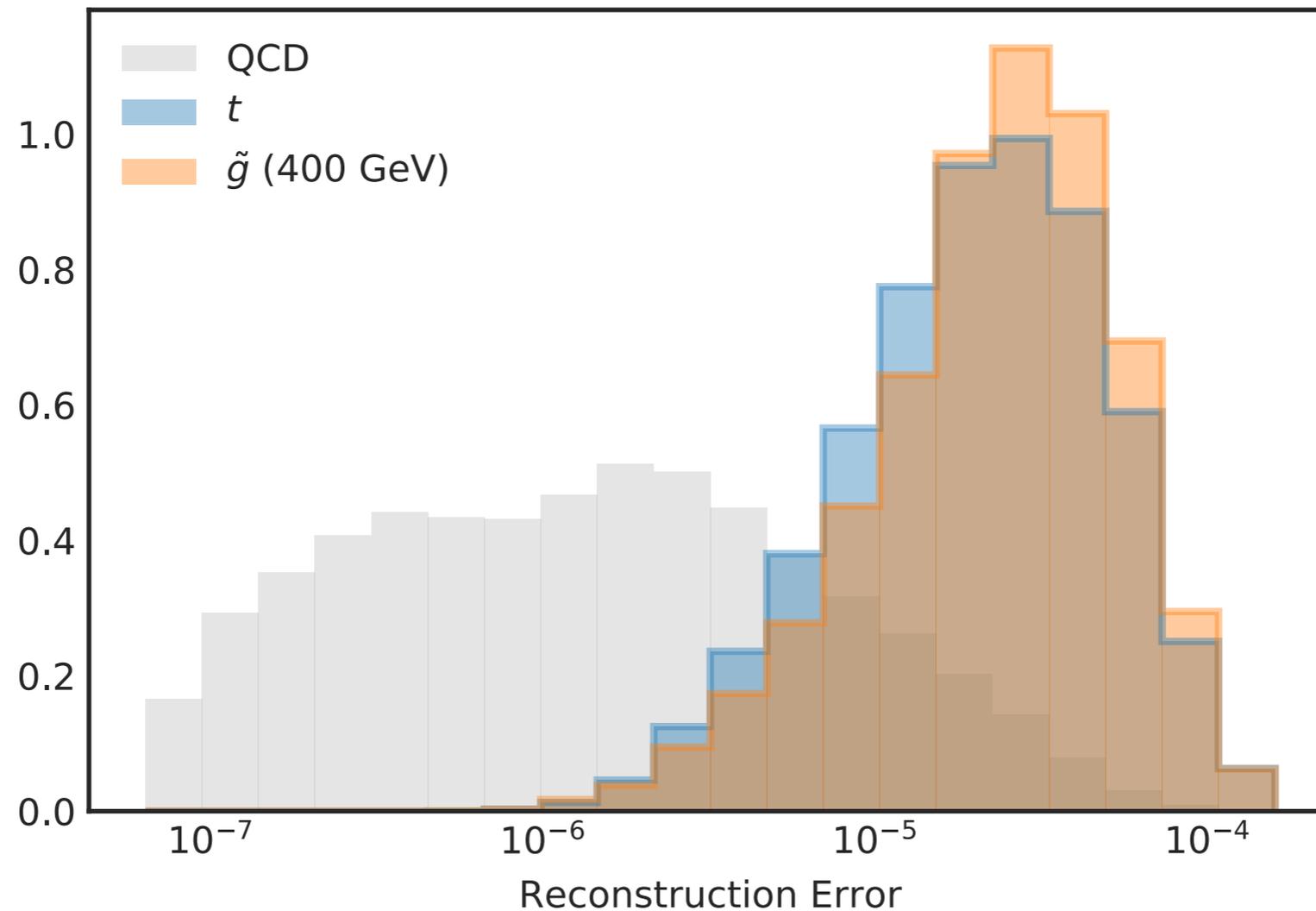
background (QCD)



signal (tops and 400 GeV RPV gluinos)



It works as an anomaly detector!



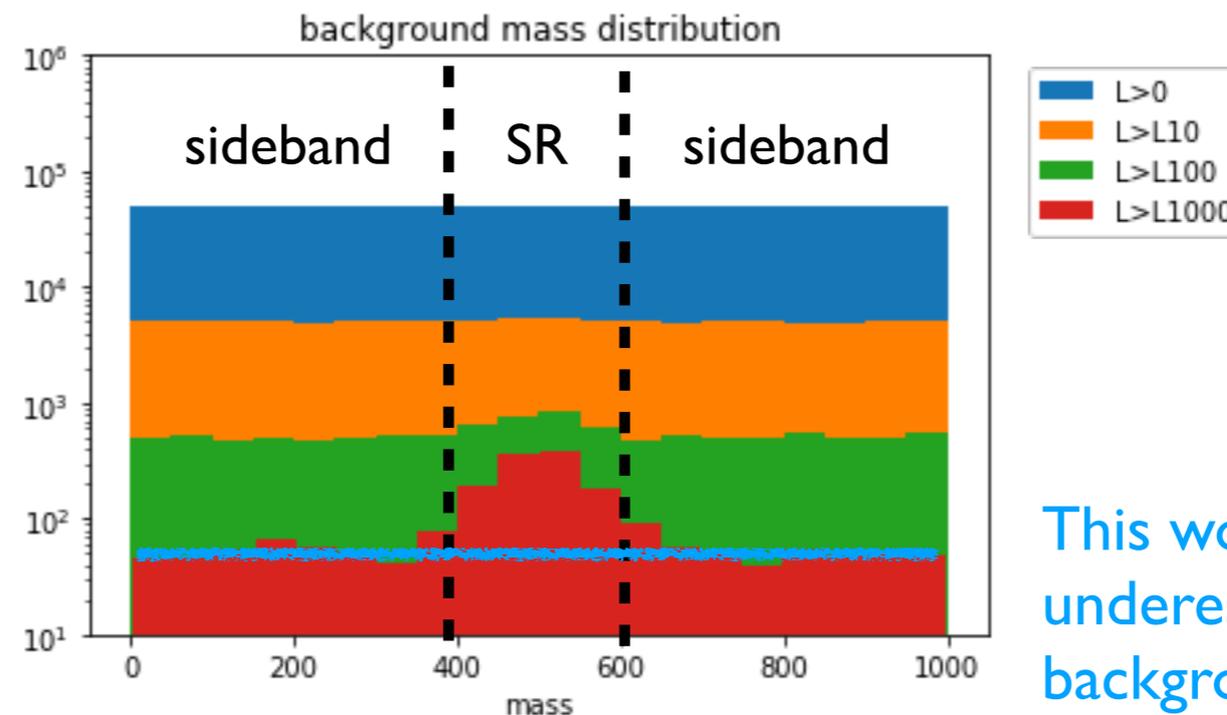
Robust against contamination with signal — can use in fully unsupervised mode

Background estimation

Discriminant is useless without an accurate background prediction.

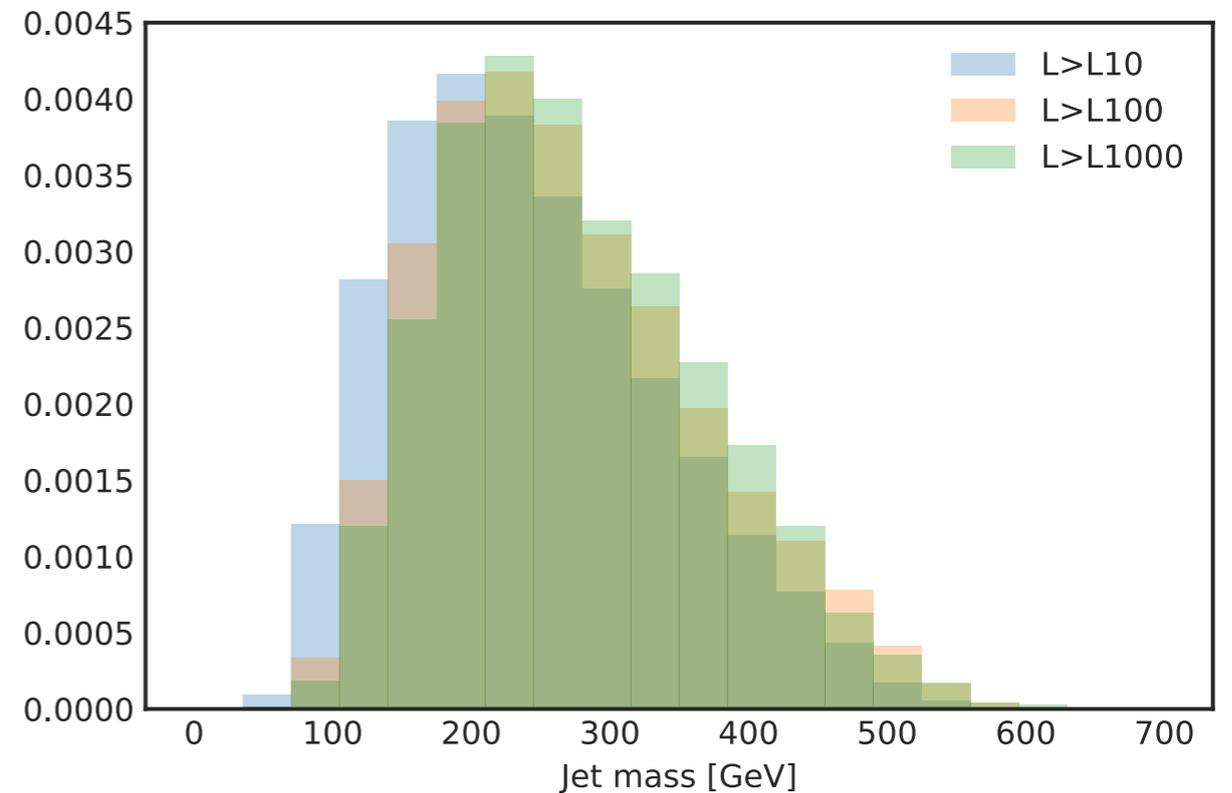
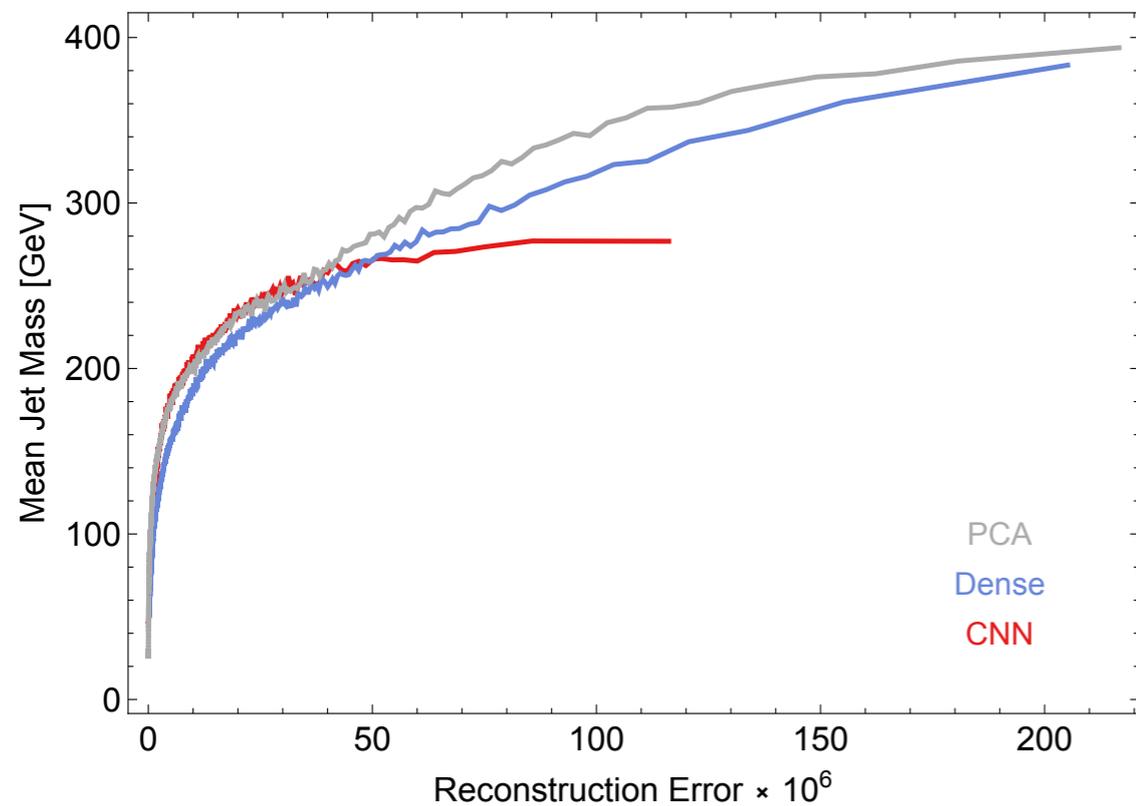
One idea: **combine autoencoder with a bump hunt in jet mass.**
Estimate backgrounds using sidebands in mass.

Only works if cutting on reconstruction error does not sculpt the mass distribution of the background!



This would greatly underestimate the background in the SR

Bump hunt with autoencoder



We found empirically that the background jet mass distribution is fairly stable against cuts on CNN AE reconstruction loss above ~ 250 GeV.

Autoencoder with explicit decorrelation

A more controlled approach to mass decorrelation would be to explicitly penalize correlations in the training of the autoencoder.

One promising method: autoencoder with adversarial decorrelation
(Heimel et al 1808.08979; based on 1611.01046, 1703.03507)

- Introduce a second NN, **the adversary**, that tries to predict the mass from the reconstruction loss.
- Penalize the total loss function when the adversary does well.

$$L_{adv} = \sum_i (f_{adv}(L_{AE}(x_i)) - m_i)^2$$

$$L_{tot} = L_{AE} - \lambda L_{adv}$$

See Tilman's talk for more on this (?)

Alternatives to adversaries

Adversaries are notoriously **tricky to train** — saddle point optimization

$$\min_{\theta_{\text{clf}}} \max_{\theta_{\text{adv}}} L_{\text{clf}}(y(\theta_{\text{clf}})) - \lambda L_{\text{adv}}(y(\theta_{\text{clf}}), m; \theta_{\text{adv}})$$

Would be great if we could achieve the same performance but with a convex regularizer term

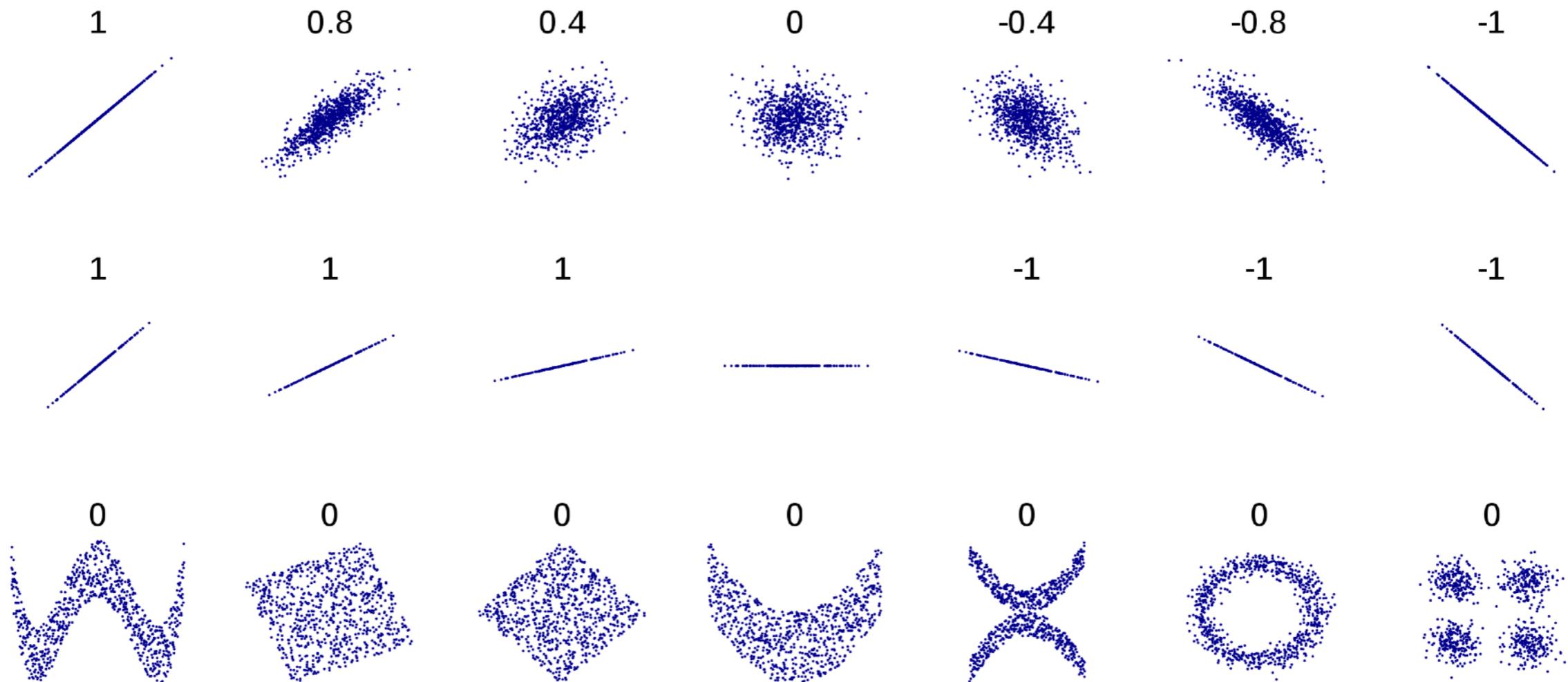
$$\min_{\theta_{\text{clf}}} L_{\text{clf}}(y(\theta_{\text{clf}})) + \lambda C_{\text{reg}}(y(\theta_{\text{clf}}), m)$$

First idea: can we just use Pearson correlation coefficient?

$$C_{\text{reg}} = R(y, m) \propto \sum_i y_i m_i$$

Problem: this only measures linear correlations

Pearson correlation



y and m can be highly correlated yet $R=0$

Distance correlation (“DisCo”)

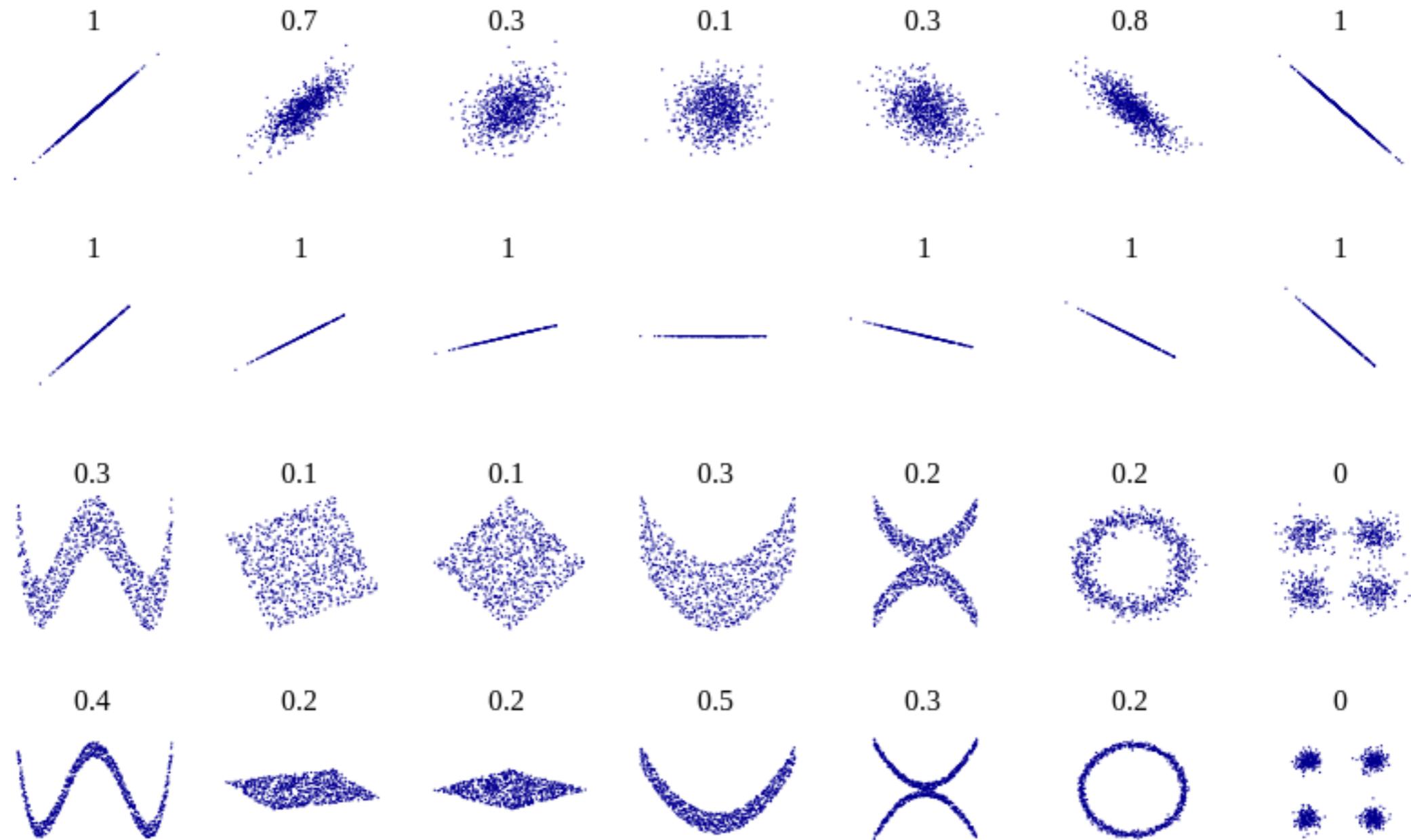
Promising idea: “distance correlation”

(Szekely, Rizzo, Bakirov 2007; Szekely & Rizzo 2009)

$$d\text{Cov}^2(X, Y) = \langle |X - X'| |Y - Y'| \rangle + \langle |X - X'| \rangle \langle |Y - Y'| \rangle - 2 \langle |X - X'| |Y - Y''| \rangle$$

- Zero iff X, Y are independent; positive otherwise
- Computationally tractable
- Straightforward sample definition — doesn't require binning

Distance correlation (“DisCo”)



Disco is sensitive to nonlinear correlations!

State of the art: ATLAS study of various decorrelation methods in context of boosted W-tagging.

Analytical / single-variable

Multivariate (MVA)

ATLAS Simulation Preliminary

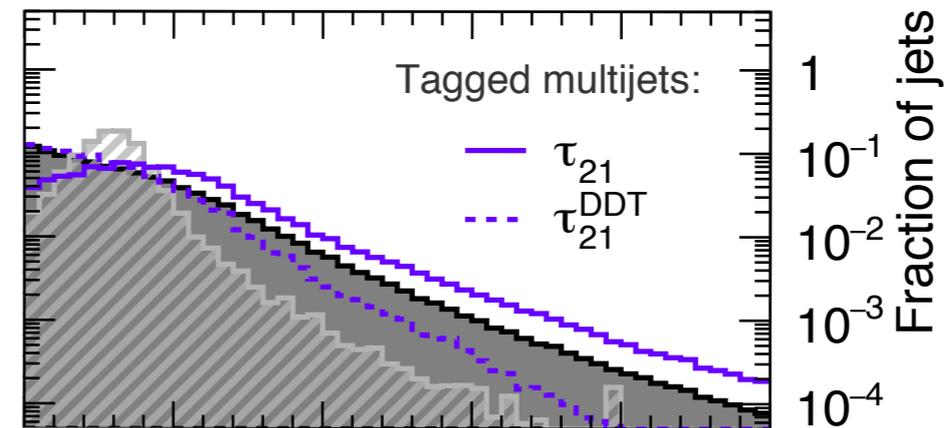
$\sqrt{s} = 13$ TeV, W jet tagging

Cuts at $\epsilon_{\text{sig}}^{\text{rel}} = 50\%$

Inclusive selection:

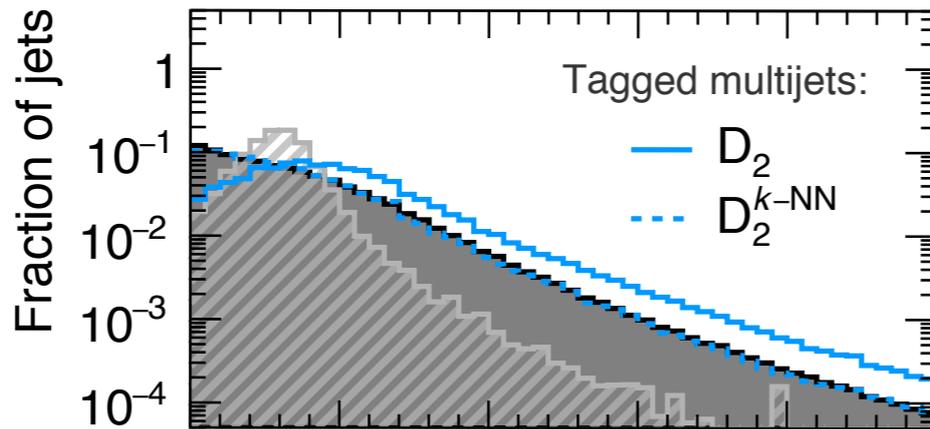
■ Multijets ▨ W jets

ATL-PHYS-PUB-2018-014

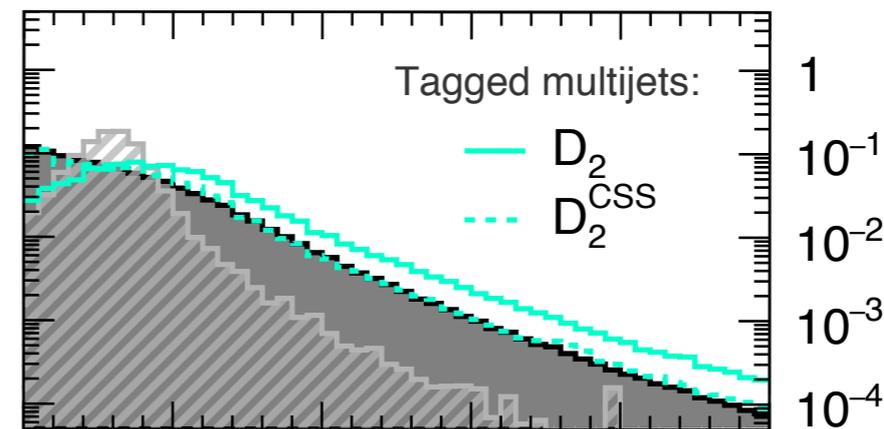


DDT

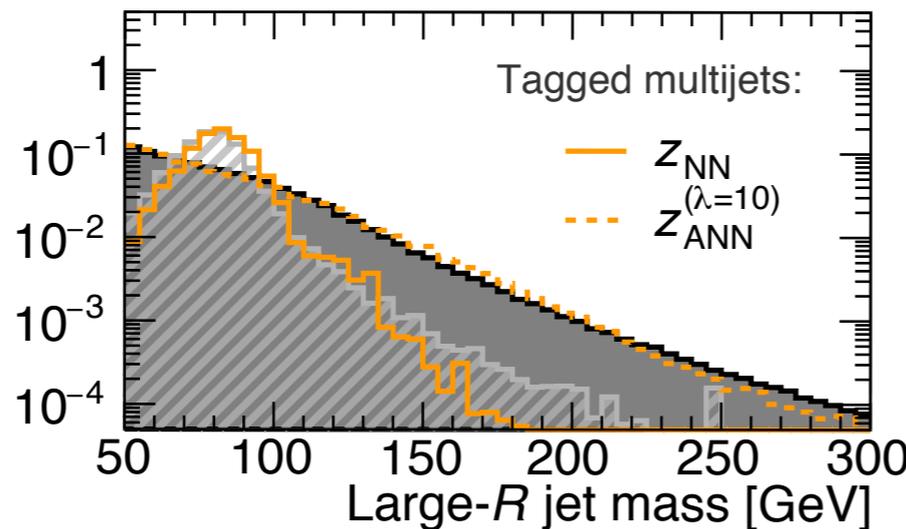
k-NN



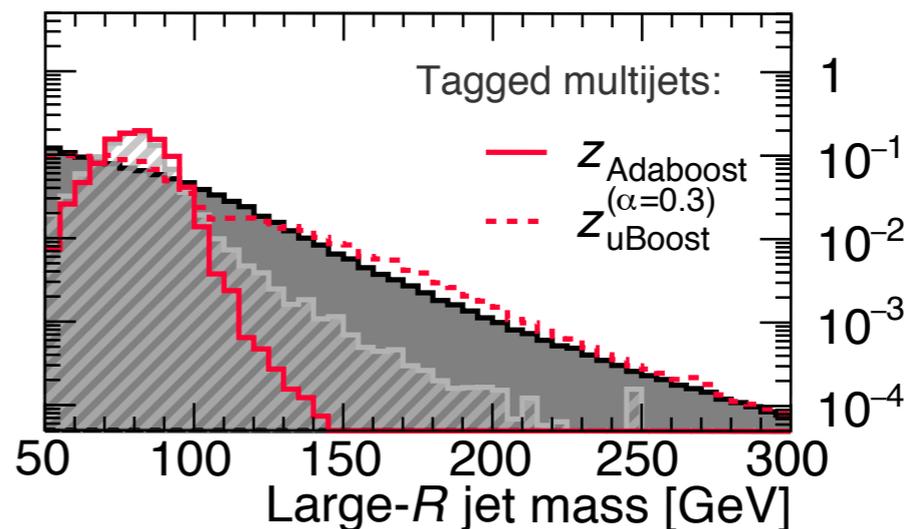
CSS



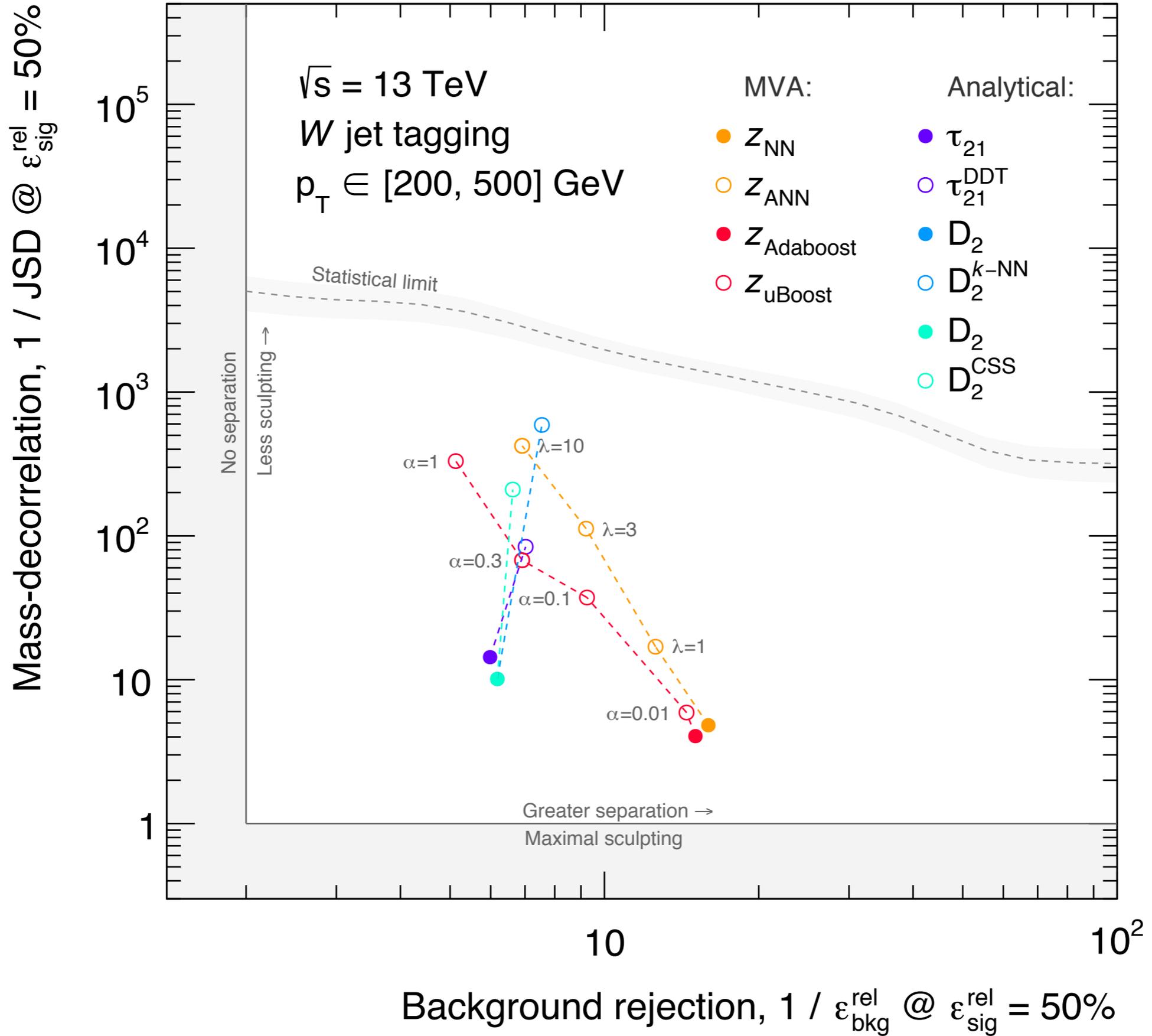
ANN



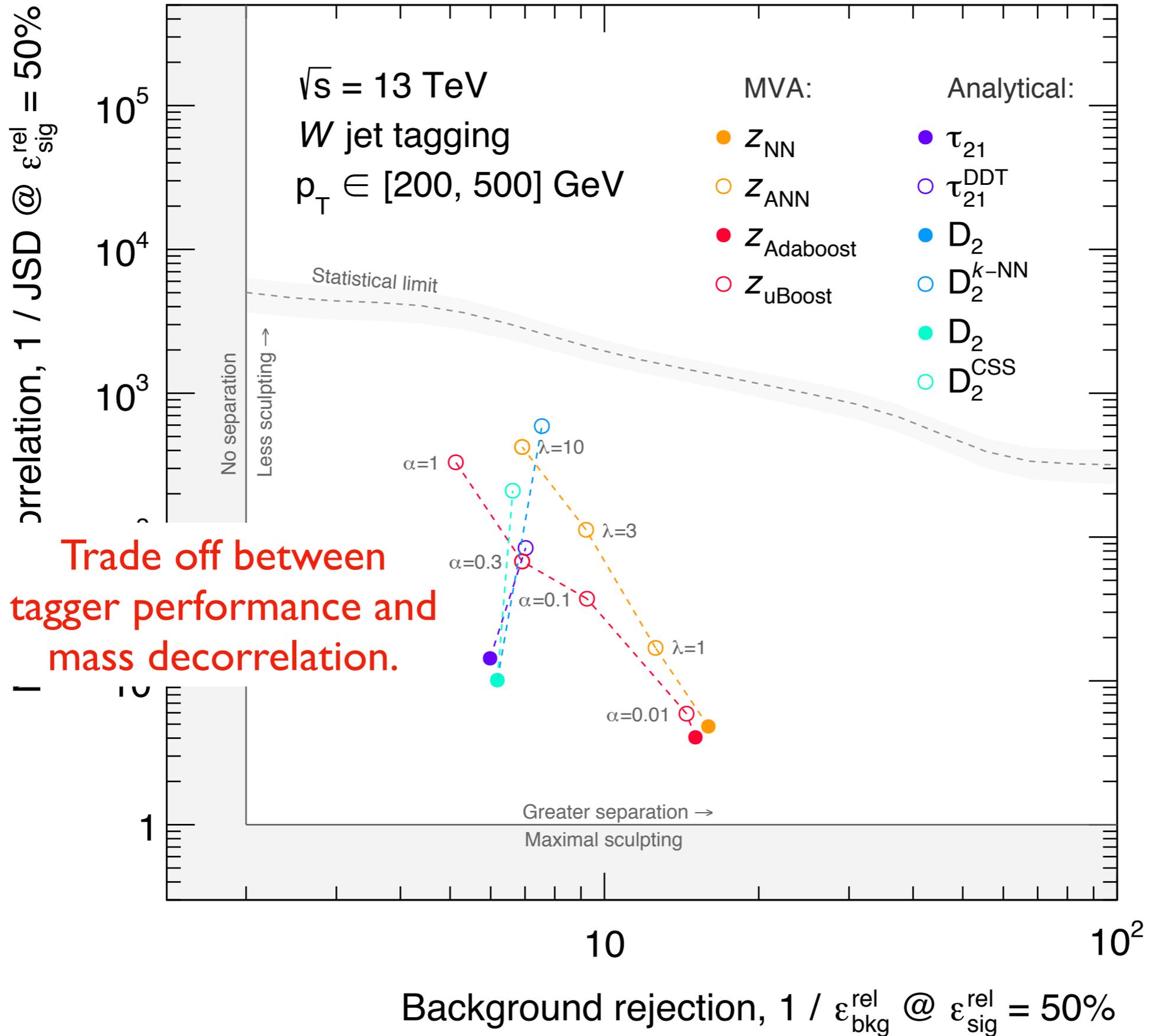
uBoost



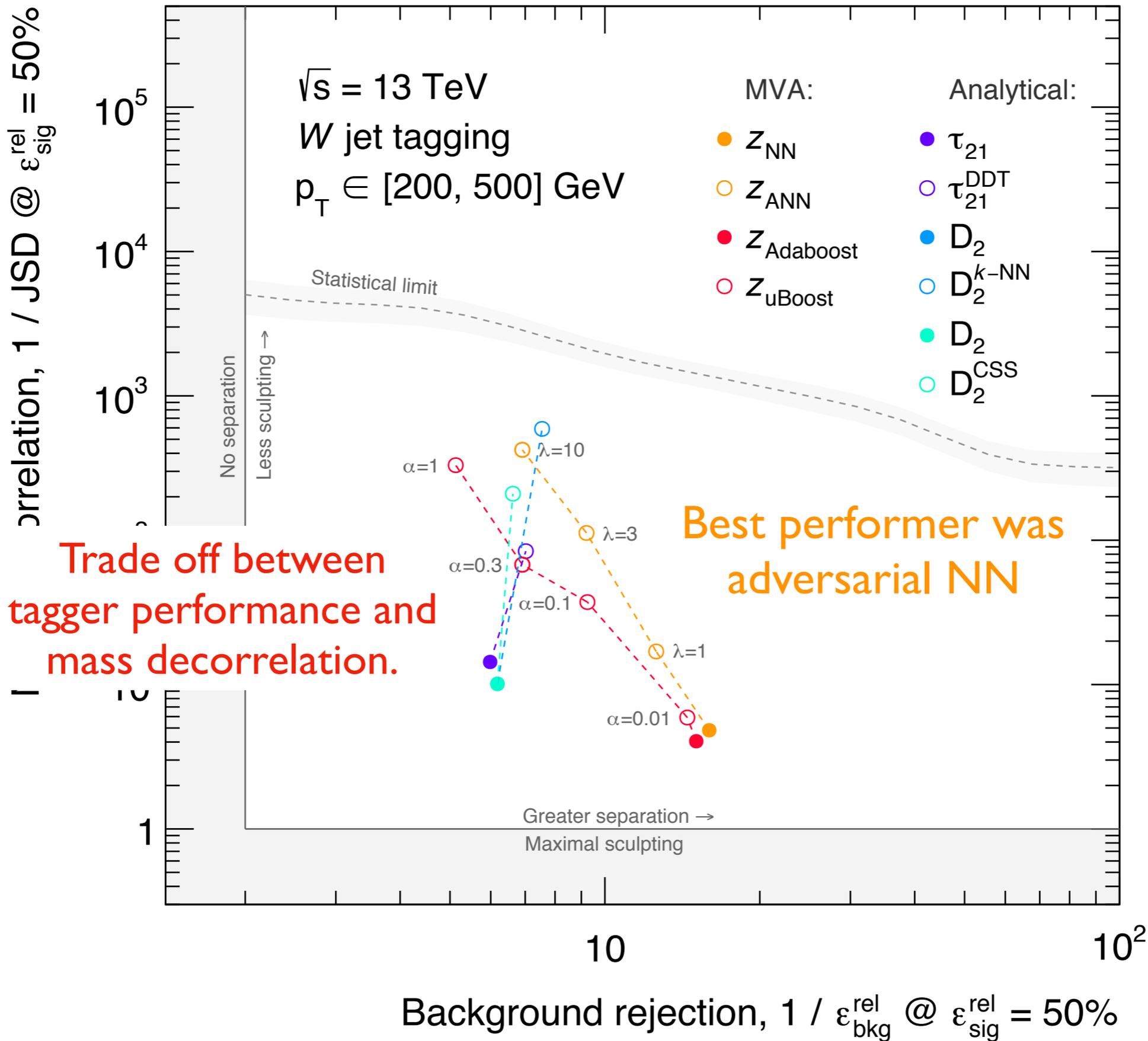
ATLAS Simulation Preliminary



ATLAS Simulation Preliminary

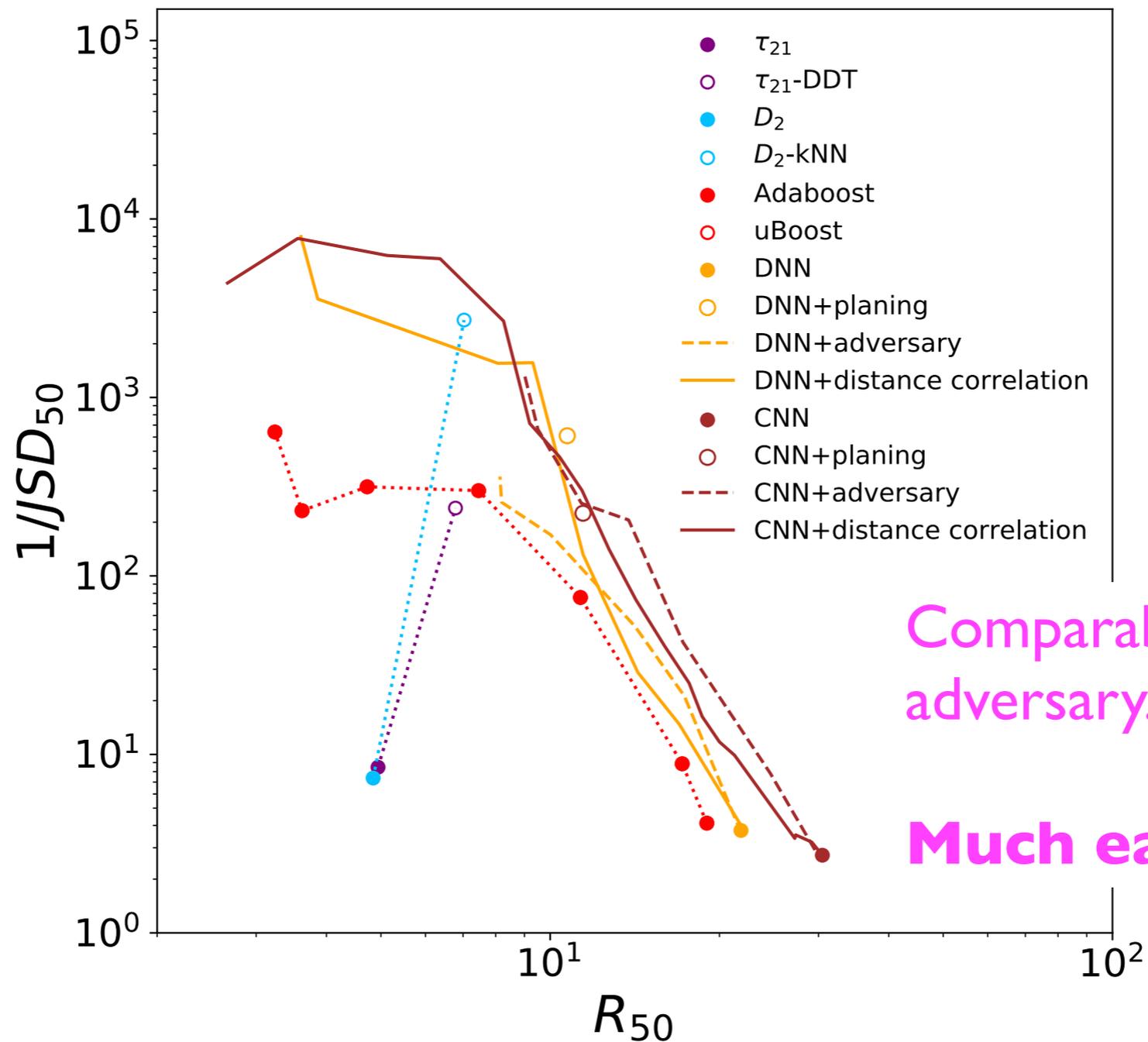
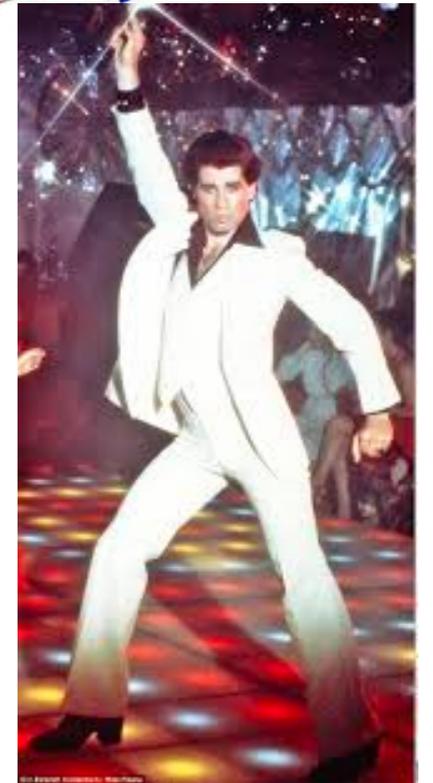
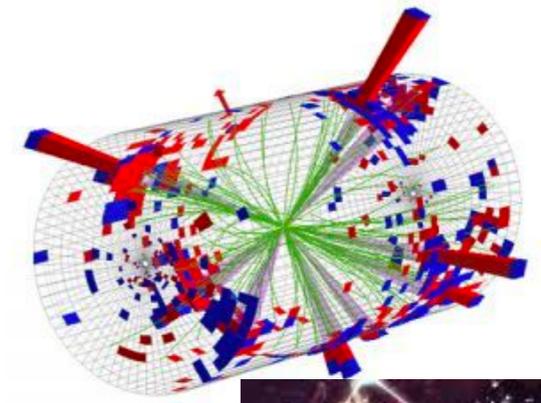


ATLAS Simulation Preliminary



DisCo decorrelation

Gregor Kasieczka & DS 2001.05310



Comparable performance to adversary.

Much easier to train.

Enhancing the bump hunt

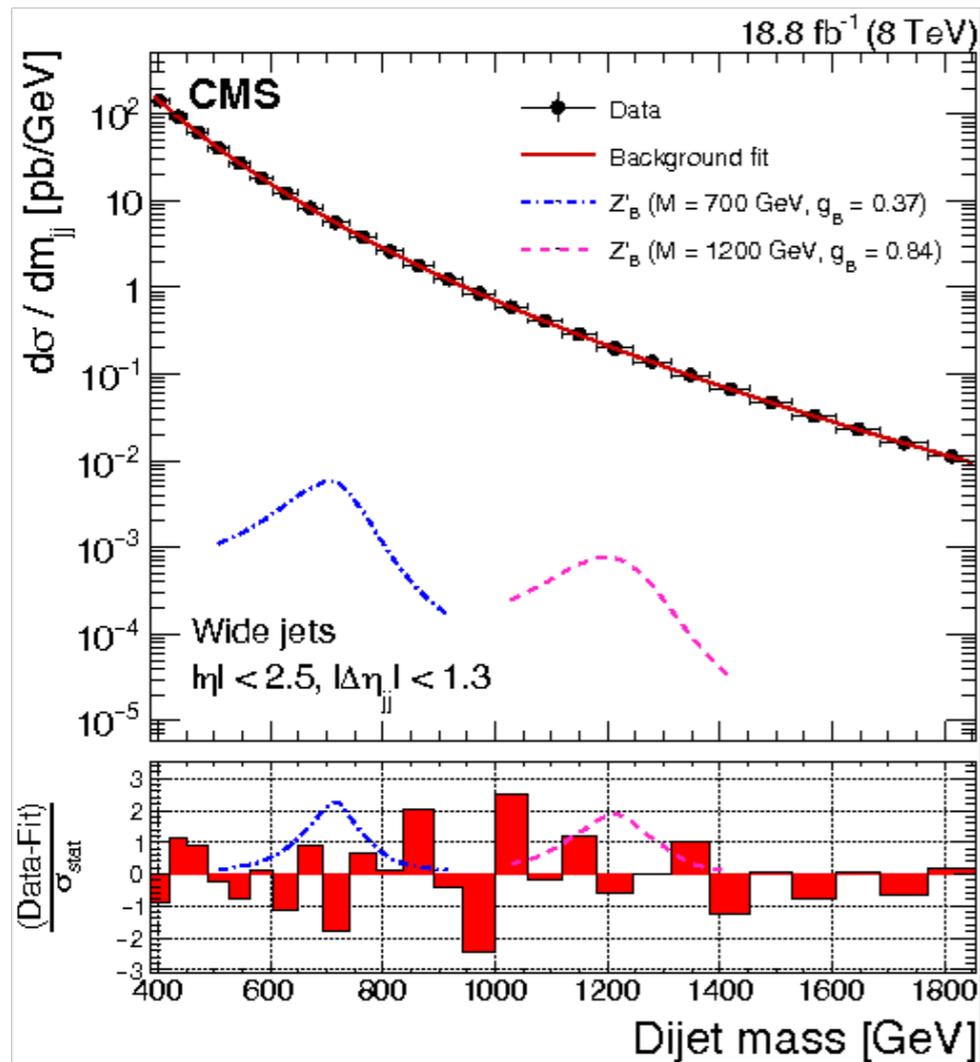
We have seen that one way to turn an autoencoder into an actual NP search is to combine it with a bump hunt.

However, what the autoencoder finds is rather uncontrolled and there is no guarantee of optimality (even asymptotically).

Can we get more if we build in the bump hunt assumption from the outset?

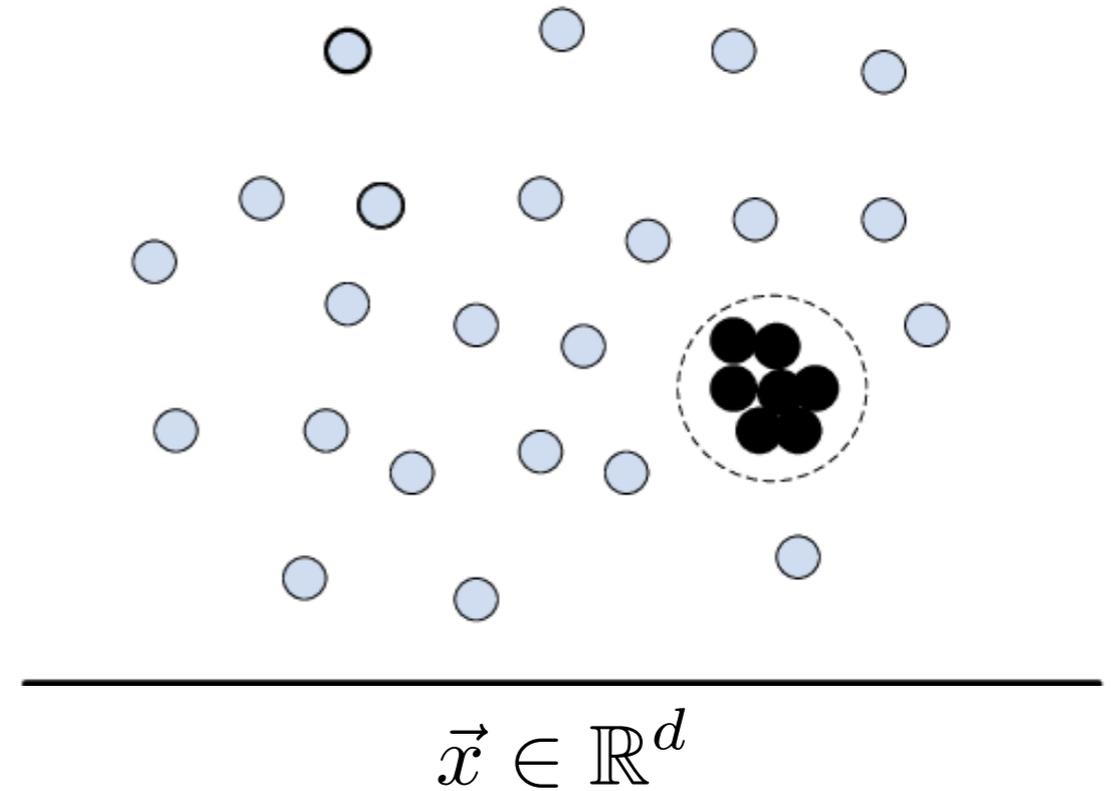
Enhancing the bump hunt

A growing number of methods aim to enhance the bump hunt using additional features:



primary resonant feature (m)

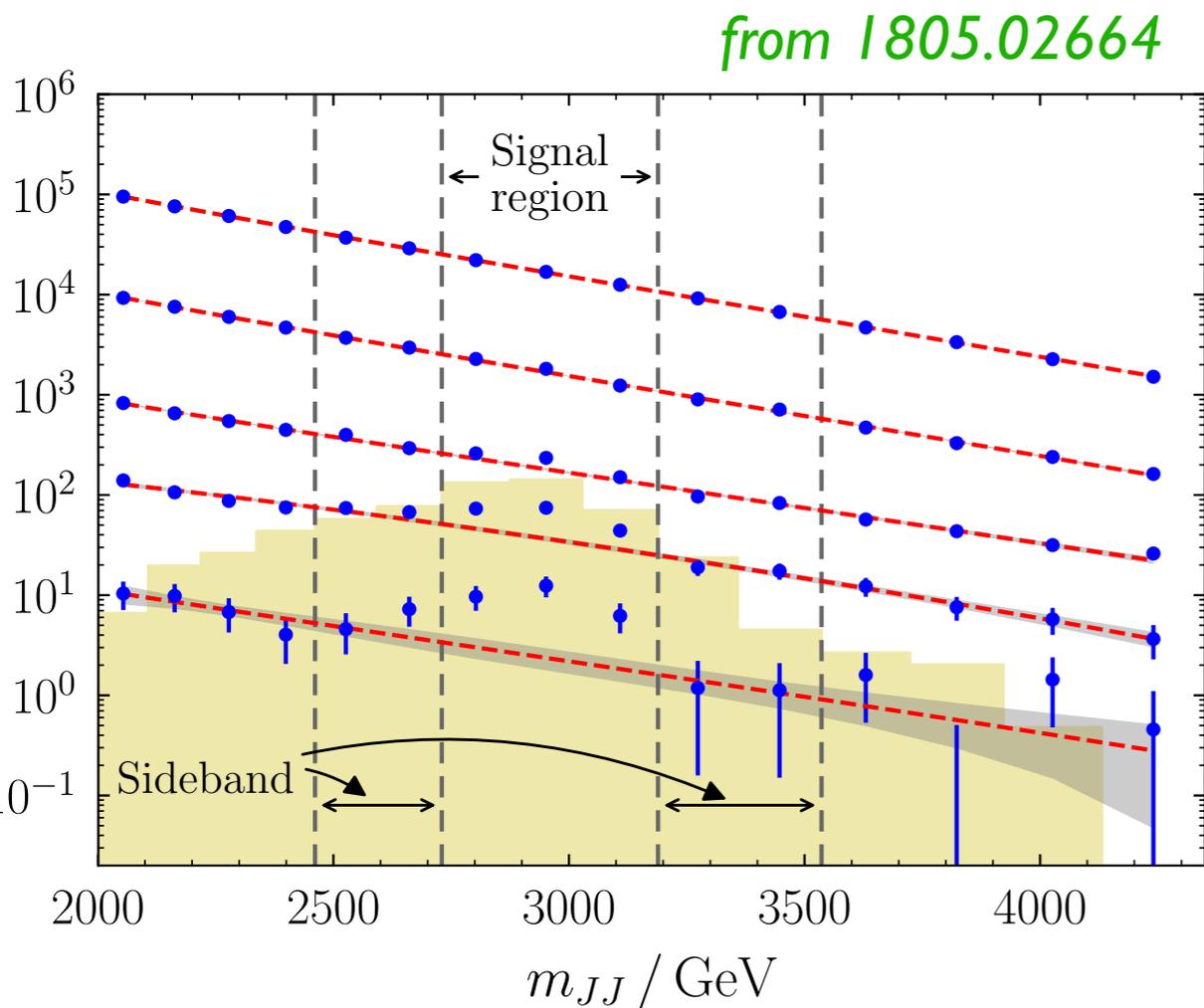
×



some additional features

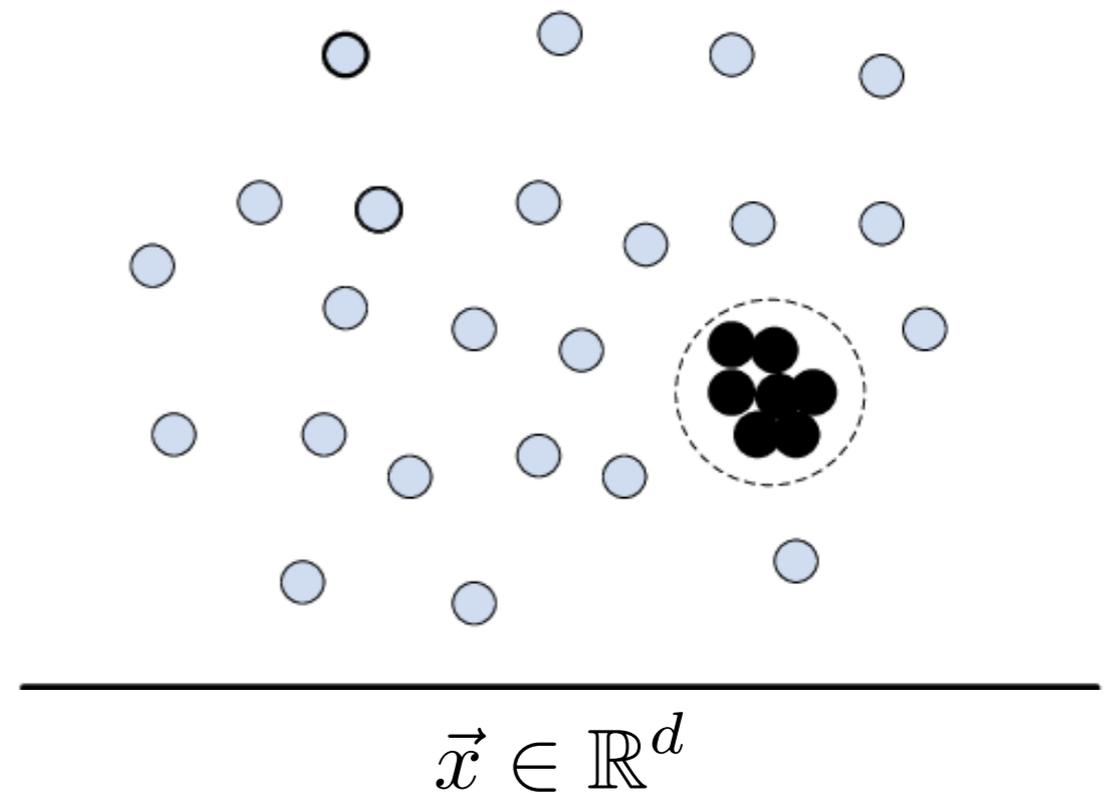
Enhancing the bump hunt

A growing number of methods aim to enhance the bump hunt using additional features:



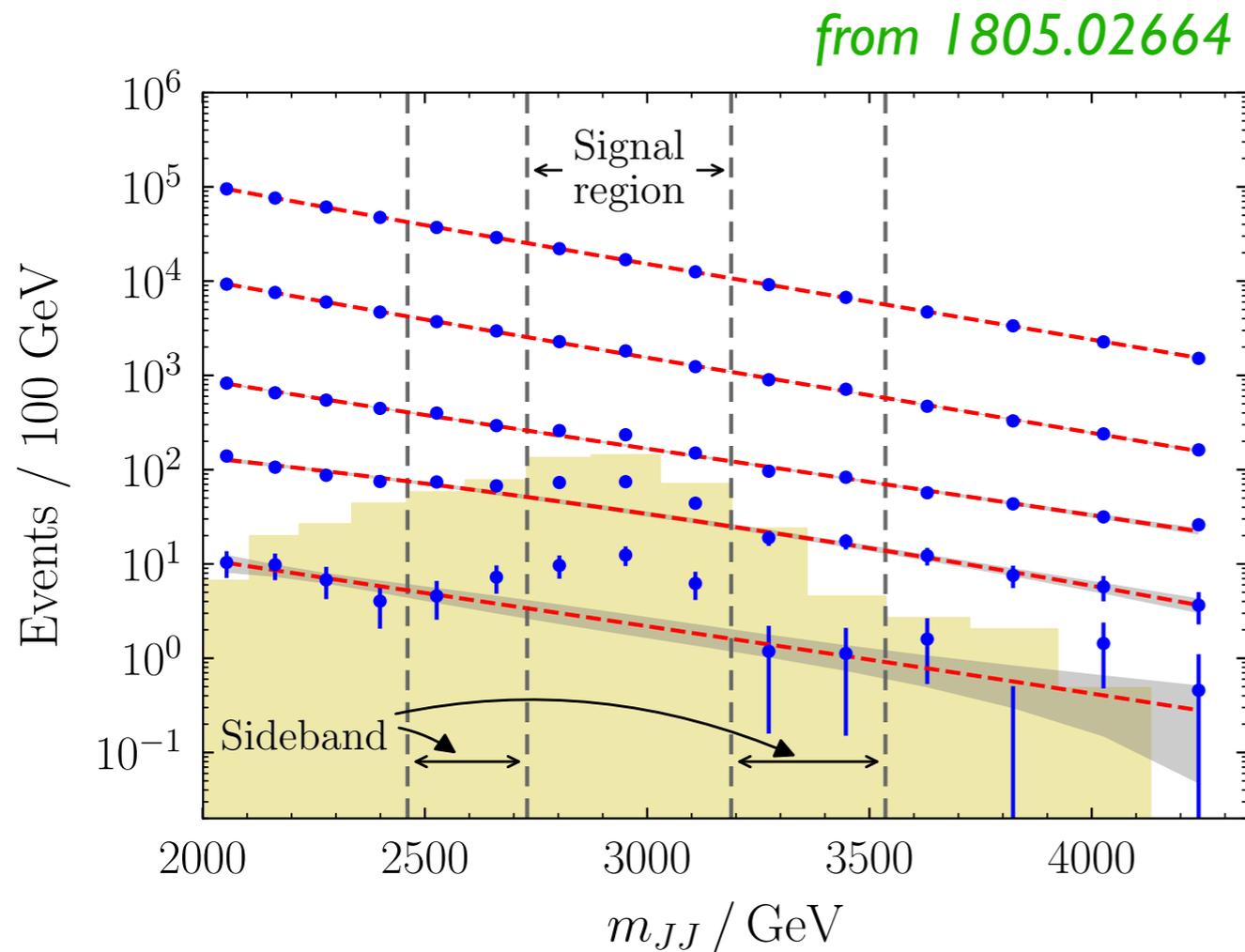
primary resonant feature (m)

×



some additional features

Enhancing the bump hunt



Use deep learning to derive something approaching the **multidimensional** likelihood ratio, directly from the data

$$R(\vec{x}) = \frac{\mathcal{L}(\vec{x} | \text{data}; m \in SR)}{\mathcal{L}(\vec{x} | B_{\text{data}}; m \in SR)}$$

Cut on $R > R_c$, enhance the bump hunt!

Enhancing the bump hunt

3 new ideas

I. CWoLa Hunting

Collins, Howe & Nachman 1805.02664, 1902.02634

Assume \vec{x} and m are completely independent

Enhancing the bump hunt

3 new ideas

I. CWoLa Hunting

Collins, Howe & Nachman 1805.02664, 1902.02634

Assume \vec{x} and m are completely independent

Train a classifier on \vec{x} to distinguish *signal region* and *sideband*. If the classifier is near-optimal, it will approach likelihood ratio

Enhancing the bump hunt

3 new ideas

I. CWoLa Hunting

Collins, Howe & Nachman 1805.02664, 1902.02634

Assume \vec{x} and m are completely independent

Train a classifier on \vec{x} to distinguish *signal region* and *sideband*. If the classifier is near-optimal, it will approach likelihood ratio

$$R_{CWoLa}(\vec{x}) = \frac{\mathcal{L}(\vec{x} | data; m \in SR)}{\mathcal{L}(\vec{x} | data; m \notin SR)}$$

Enhancing the bump hunt

3 new ideas

I. CWoLa Hunting

Collins, Howe & Nachman 1805.02664, 1902.02634

Assume \vec{x} and m are completely independent

Train a classifier on \vec{x} to distinguish *signal region* and *sideband*. If the classifier is near-optimal, it will approach likelihood ratio

$$\begin{aligned} R_{CWoLa}(\vec{x}) &= \frac{\mathcal{L}(\vec{x}|data; m \in SR)}{\mathcal{L}(\vec{x}|data; m \notin SR)} \\ &= \frac{\mathcal{L}(\vec{x}|data; m \in SR)}{\mathcal{L}(\vec{x}|B_{data}; m \notin SR)} \end{aligned}$$

Enhancing the bump hunt

3 new ideas

I. CWoLa Hunting

Collins, Howe & Nachman 1805.02664, 1902.02634

Assume \vec{x} and m are completely independent

Train a classifier on \vec{x} to distinguish *signal region* and *sideband*. If the classifier is near-optimal, it will approach likelihood ratio

$$\begin{aligned} R_{CWoLa}(\vec{x}) &= \frac{\mathcal{L}(\vec{x}|data; m \in SR)}{\mathcal{L}(\vec{x}|data; m \notin SR)} \\ &= \frac{\mathcal{L}(\vec{x}|data; m \in SR)}{\mathcal{L}(\vec{x}|B_{data}; m \notin SR)} \\ &= \frac{\mathcal{L}(\vec{x}|data; m \in SR)}{\mathcal{L}(\vec{x}|B_{data}; m \in SR)} \end{aligned}$$

Enhancing the bump hunt

3 new ideas

I. CWoLa Hunting

Collins, Howe & Nachman 1805.02664, 1902.02634

Assume \vec{x} and m are completely independent

Train a classifier on \vec{x} to distinguish *signal region* and *sideband*. If the classifier is near-optimal, it will approach likelihood ratio

$$\begin{aligned}
 R_{CWoLa}(\vec{x}) &= \frac{\mathcal{L}(\vec{x}|data; m \in SR)}{\mathcal{L}(\vec{x}|data; m \notin SR)} \\
 &= \frac{\mathcal{L}(\vec{x}|data; m \in SR)}{\mathcal{L}(\vec{x}|B_{data}; m \notin SR)} \\
 &= \frac{\mathcal{L}(\vec{x}|data; m \in SR)}{\mathcal{L}(\vec{x}|B_{data}; m \in SR)}
 \end{aligned}$$

“Classification Without Labels”

Enhancing the bump hunt

3 new ideas

2. AnoDE: Anomaly Detection with Density Estimation

Nachman & DS 2001.04990

Directly learn the conditional probability densities from the data

$$\mathcal{L}(\vec{x}|data; m \in SR) \quad \mathcal{L}(\vec{x}|data; m \notin SR) = \mathcal{L}(\vec{x}|B_{data}; m \notin SR)$$

Enhancing the bump hunt

3 new ideas

2. AnoDE: Anomaly Detection with Density Estimation

Nachman & DS 2001.04990

Directly learn the conditional probability densities from the data

$$\mathcal{L}(\vec{x}|data; m \in SR) \quad \mathcal{L}(\vec{x}|data; m \notin SR) = \mathcal{L}(\vec{x}|B_{data}; m \notin SR)$$

interpolate in (x,m)



Enhancing the bump hunt

3 new ideas

2. AnoDE: Anomaly Detection with Density Estimation

Nachman & DS 2001.04990

Directly learn the conditional probability densities from the data

$$\mathcal{L}(\vec{x}|data; m \in SR) \quad \mathcal{L}(\vec{x}|data; m \notin SR) = \mathcal{L}(\vec{x}|B_{data}; m \notin SR)$$

interpolate in (x,m)



$$\mathcal{L}(\vec{x}|B_{data}; m \in SR)$$

Enhancing the bump hunt

3 new ideas

2. AnoDE: Anomaly Detection with Density Estimation Nachman & DS 2001.04990

Directly learn the conditional probability densities from the data

$$\mathcal{L}(\vec{x}|data; m \in SR) \quad \mathcal{L}(\vec{x}|data; m \notin SR) = \mathcal{L}(\vec{x}|B_{data}; m \notin SR)$$

interpolate in (x,m)



$$\mathcal{L}(\vec{x}|B_{data}; m \in SR)$$

Construct the likelihood ratio: $R(\vec{x}) = \frac{\mathcal{L}(\vec{x}|data; m \in SR)}{\mathcal{L}(\vec{x}|B_{data}; m \in SR)}$

Enhancing the bump hunt

3 new ideas

2. AnoDE: Anomaly Detection with Density Estimation Nachman & DS 2001.04990

Directly learn the conditional probability densities from the data

$$\mathcal{L}(\vec{x}|data; m \in SR) \quad \mathcal{L}(\vec{x}|data; m \notin SR) = \mathcal{L}(\vec{x}|B_{data}; m \notin SR)$$

interpolate in (x,m)



$$\mathcal{L}(\vec{x}|B_{data}; m \in SR)$$

Construct the likelihood ratio:
$$R(\vec{x}) = \frac{\mathcal{L}(\vec{x}|data; m \in SR)}{\mathcal{L}(\vec{x}|B_{data}; m \in SR)}$$

Recent breakthroughs in (neural) density estimation make this possible in high dimensions. We used conditional MAF (Papamakarios et al 1705.07057) but many other density estimators are possible

Enhancing the bump hunt

3 new ideas

3. SALAD: Simulation Assisted Likelihood-free Anomaly Detection Andreassen, Nachman & DS 2001.05001

Simulations in HEP are very good but not good enough to directly compare against data. If we could reweight simulations to match data they would be much more useful!

- Train a classifier on data vs simulation in sidebands. If this classifier is near-optimal, it will approach the likelihood ratio (1907.08209)

$$w(\vec{x}) = \frac{\mathcal{L}(\vec{x}|B_{data}, m \notin SR)}{\mathcal{L}(\vec{x}|B_{sim}, m \notin SR)}$$

- Interpolate into SR
- Using reweighted simulation, generate a sample that follows $\mathcal{L}(\vec{x}|B_{data}, m \in SR)$
- Train a classifier to distinguish data from this sample
- Obtain a discriminant that approaches

Enhancing the bump hunt

3 new ideas

3. SALAD: Simulation Assisted Likelihood-free Anomaly Detection Andreassen, Nachman & DS 2001.05001

Simulations in HEP are very good but not good enough to directly compare against data. If we could reweight simulations to match data they would be much more useful!

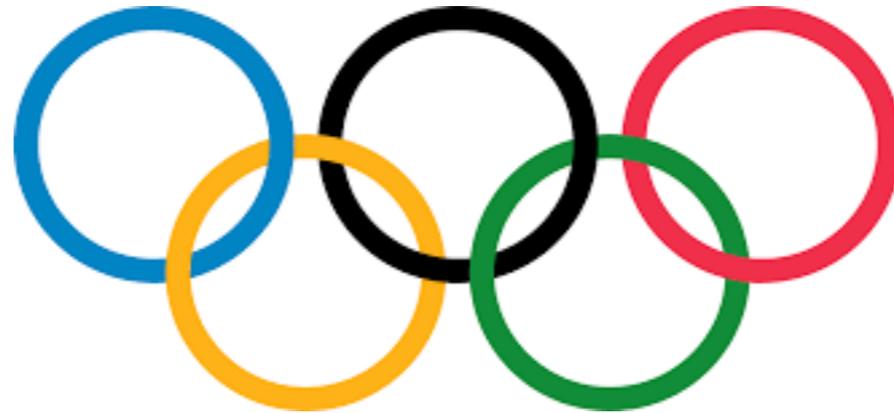
- Train a classifier on data vs simulation in sidebands. If this classifier is near-optimal, it will approach the likelihood ratio (1907.08209)

$$w(\vec{x}) = \frac{\mathcal{L}(\vec{x}|B_{data}, m \notin SR)}{\mathcal{L}(\vec{x}|B_{sim}, m \notin SR)}$$

- Interpolate into SR
- Using reweighted simulation, generate a sample that follows $\mathcal{L}(\vec{x}|B_{data}, m \in SR)$
- Train a classifier to distinguish data from this sample
- Obtain a discriminant that approaches $R(\vec{x}) = \frac{\mathcal{L}(\vec{x}|data; m \in SR)}{\mathcal{L}(\vec{x}|B_{data}; m \in SR)}$

LHC Olympics 2020

To facilitate a meaningful comparison between different approaches and to spur the development of new ones, in April 2019, **Gregor Kasieczka, Ben Nachman and I** initiated an anomaly detection data challenge:



The LHC Olympics 2020

<https://indico.cern.ch/event/809820/page/19002-lhcolympics2020>

LHC Olympics 2020: Black Boxes

We prepared three black boxes of simulated data:

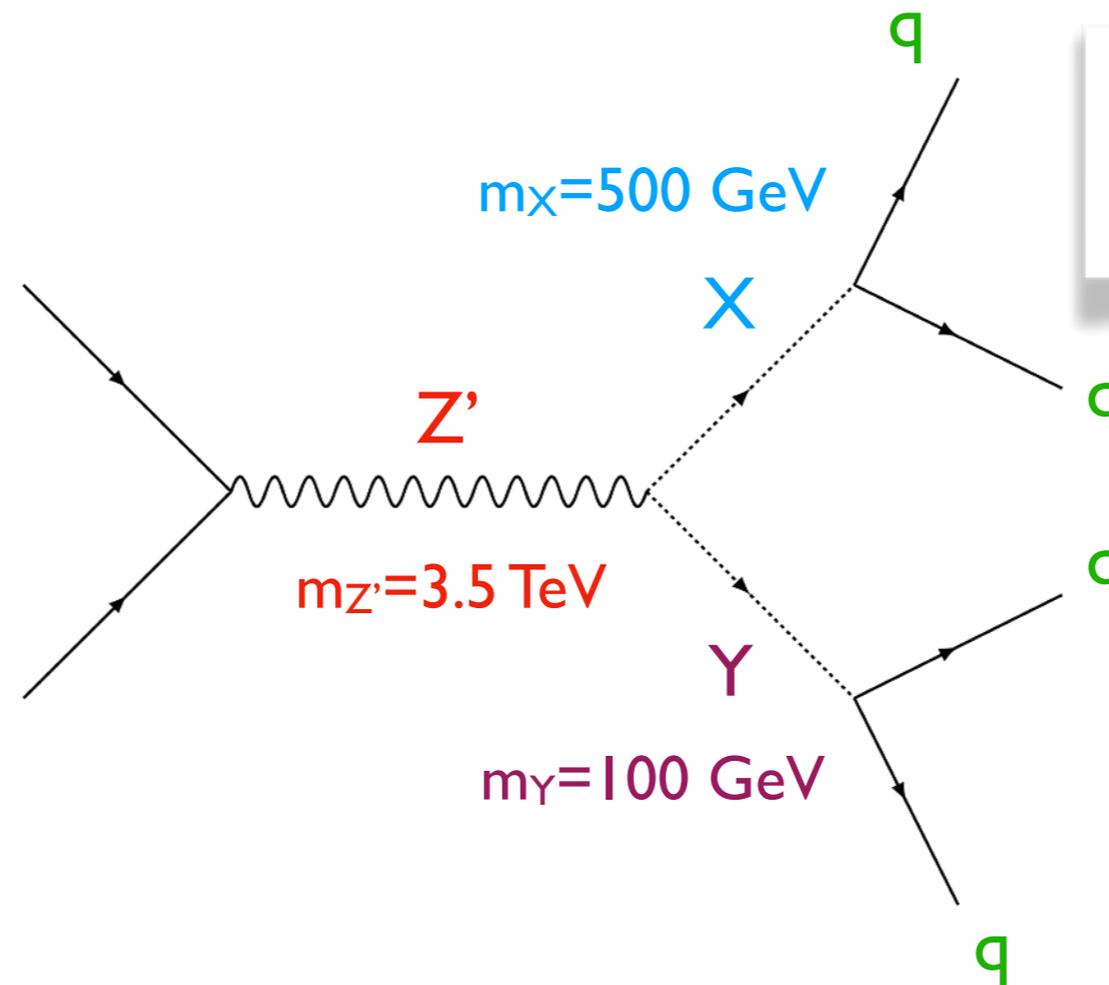
- 1 million events each
- 4-vectors of every reconstructed particle (hadron) in the event
- Particle ID, charge, etc not included
- Single R=1 jet trigger $p_T > 1.2$ TeV
- Black boxes are meant to be representative of actual data, meaning they are mostly background and may contain signals of new physics

In addition, a sample of 1M QCD dijet events (produced with Pythia8 and Delphes3.4.1) was provided as a background sample.

<https://doi.org/10.5281/zenodo.3547721>

LHC Olympics 2020: R&D Dataset

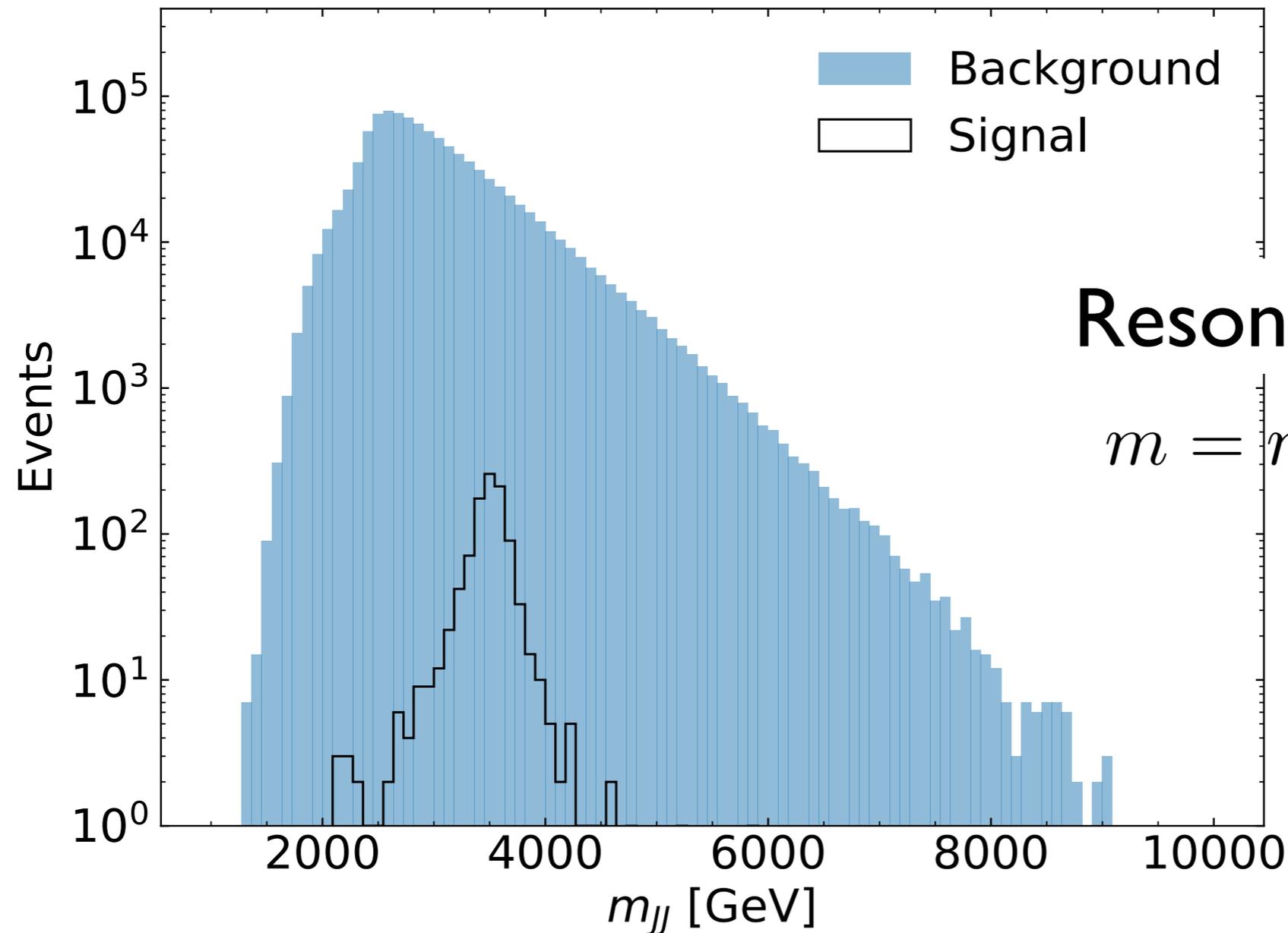
Prior to the challenge, we also released a labeled R&D dataset consisting of 1M QCD dijet events and 100k signal events



No explicit search at the LHC for this scenario!

<https://doi.org/10.5281/zenodo.2629072>

LHC Olympics 2020: R&D Dataset



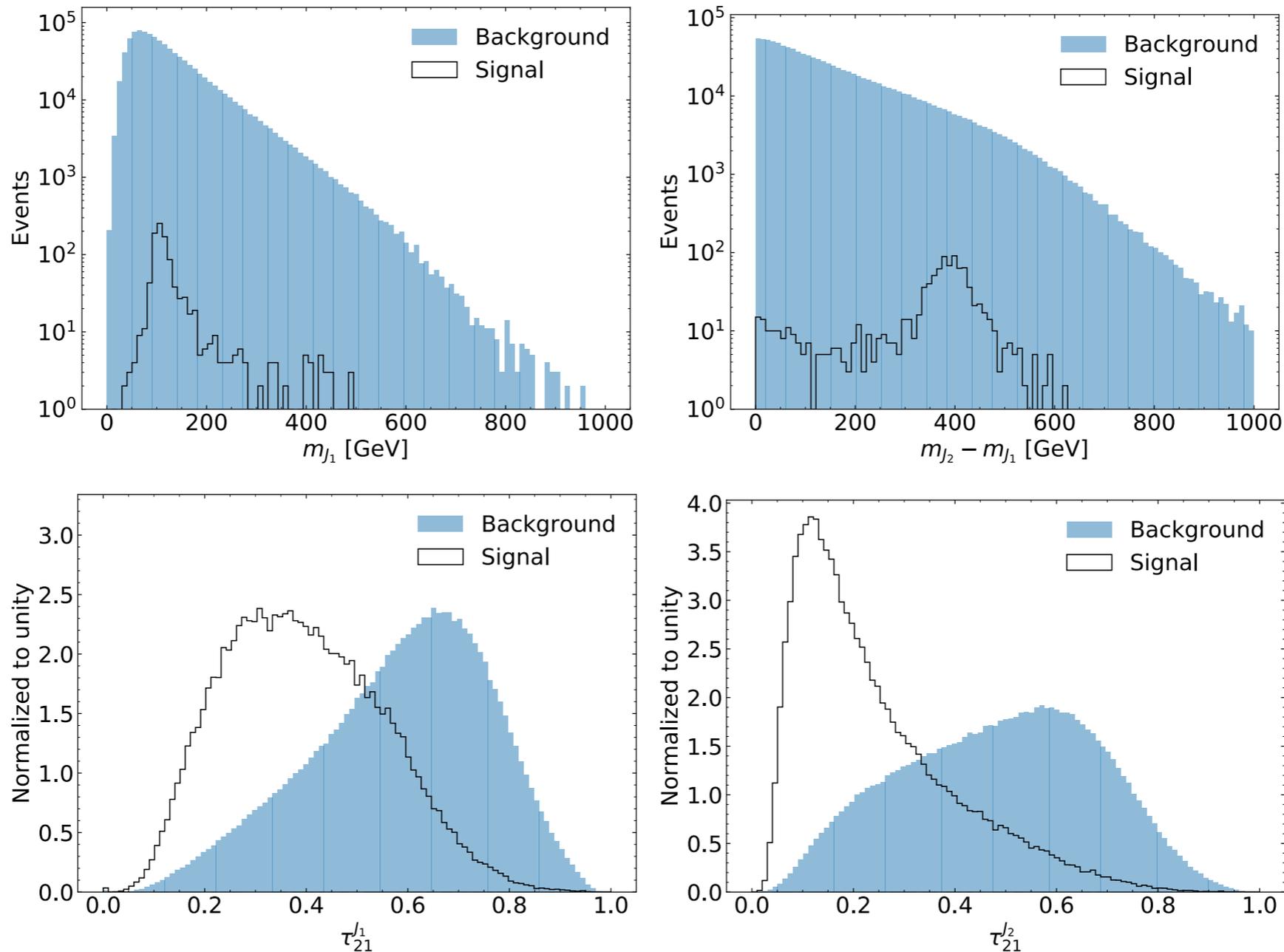
Resonant feature

$$m = m_{Z'} = m_{JJ}$$

$S=500, B=500,000, B_{SR}=61,000$

$S/B_{SR} \sim 6 \times 10^{-3}, S/\sqrt{B_{SR}} \sim 1.5$

LHC Olympics 2020: R&D Dataset

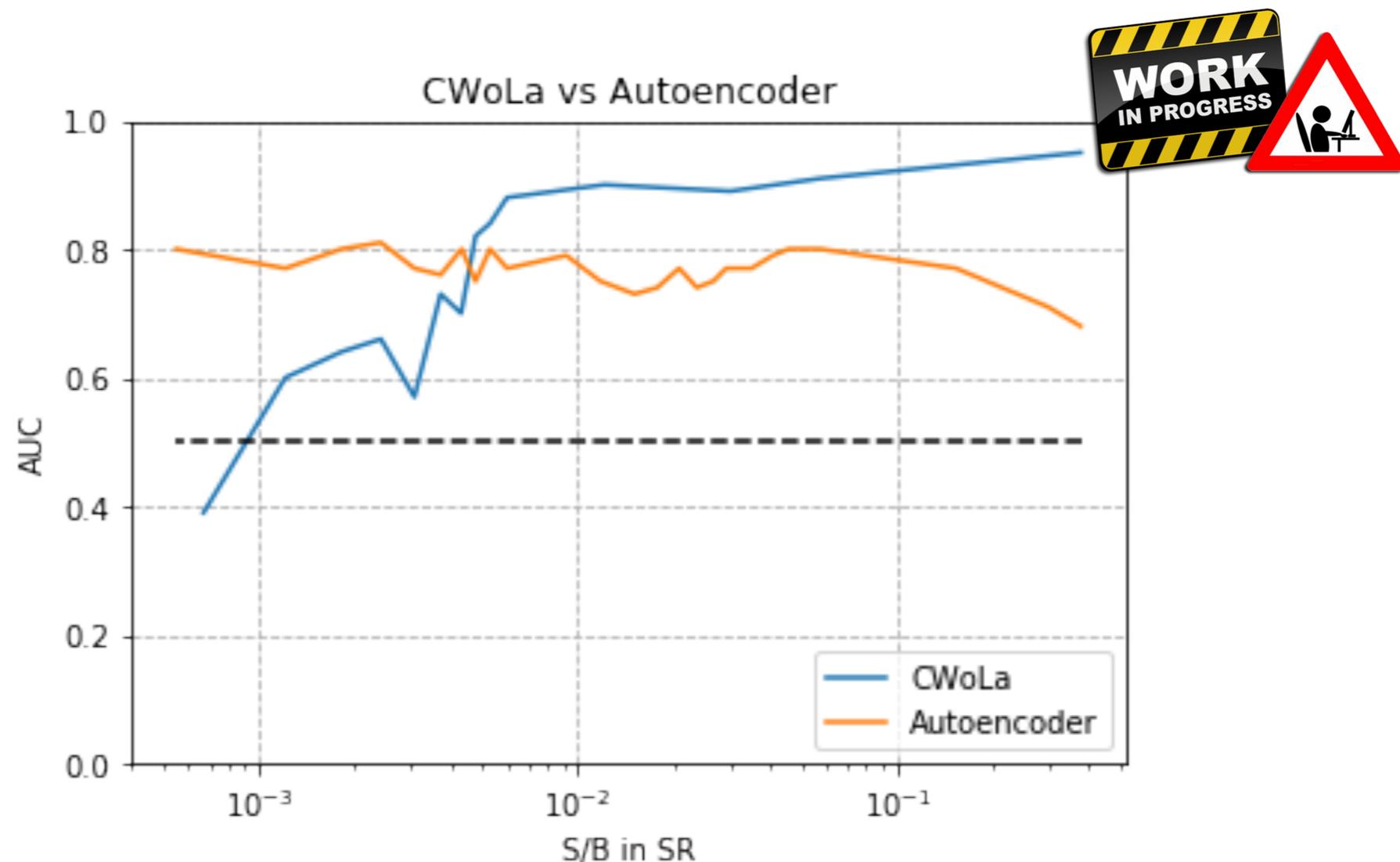


Additional features: $x = (m_{J_1}, m_{J_2}, \tau_{21}^{J_1}, \tau_{21}^{J_2})$

Comparing CWoLa vs Autoencoders with LHCO R&D Dataset

Pablo Martin, Ben Nachman & DS work in progress

- Test performance of both methods for different S/B ratios



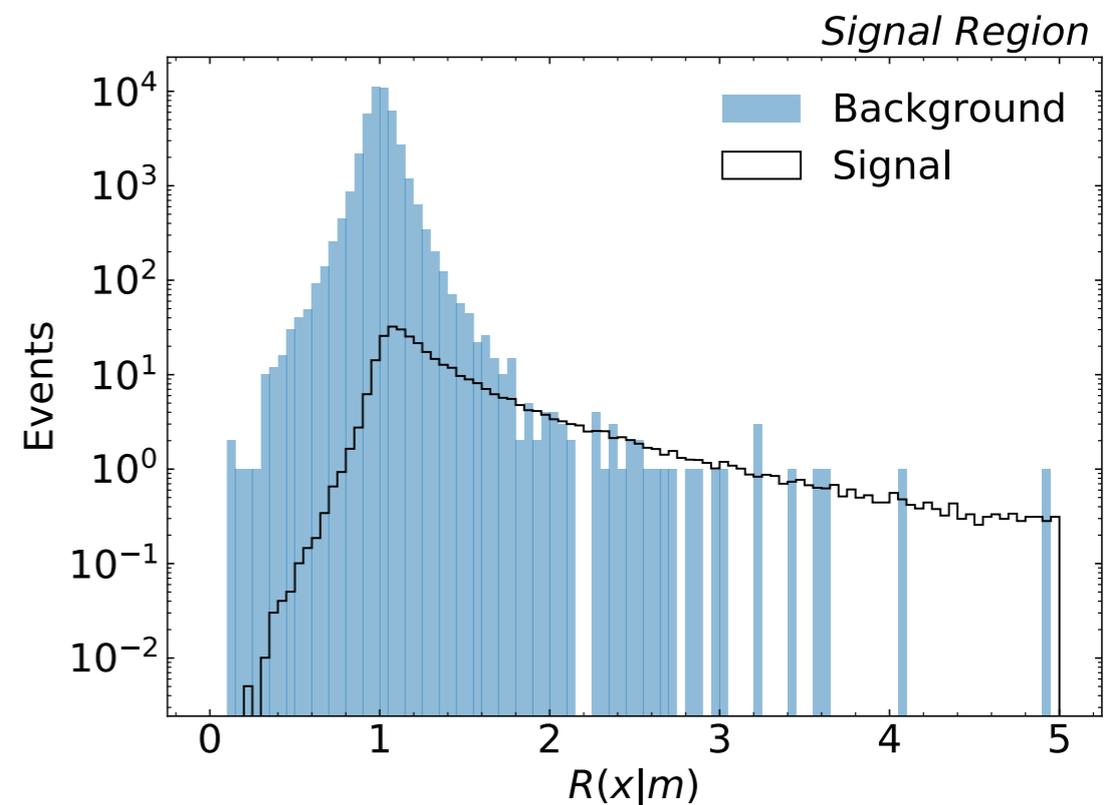
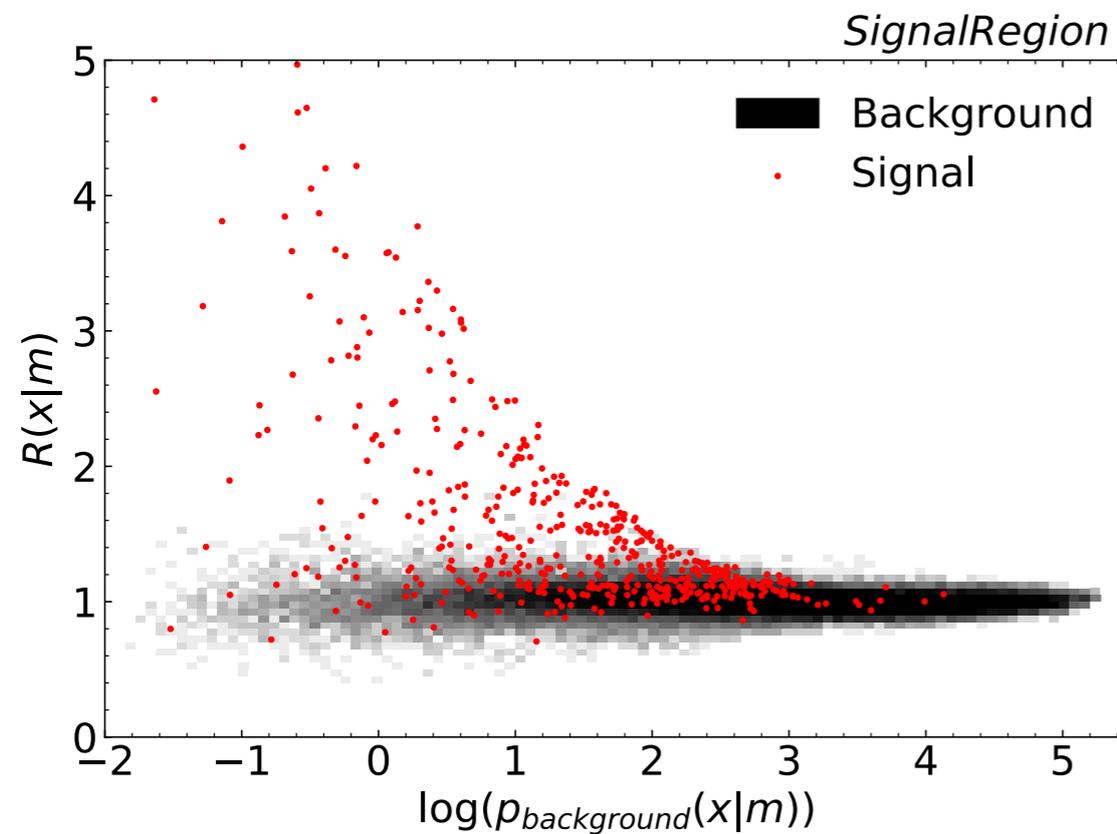
⇒ CWoLa performs better at large cross sections

⇒ Autoencoder solid at very low cross sections

Complementary techniques!

ANODE: Results on LHCO R&D Dataset

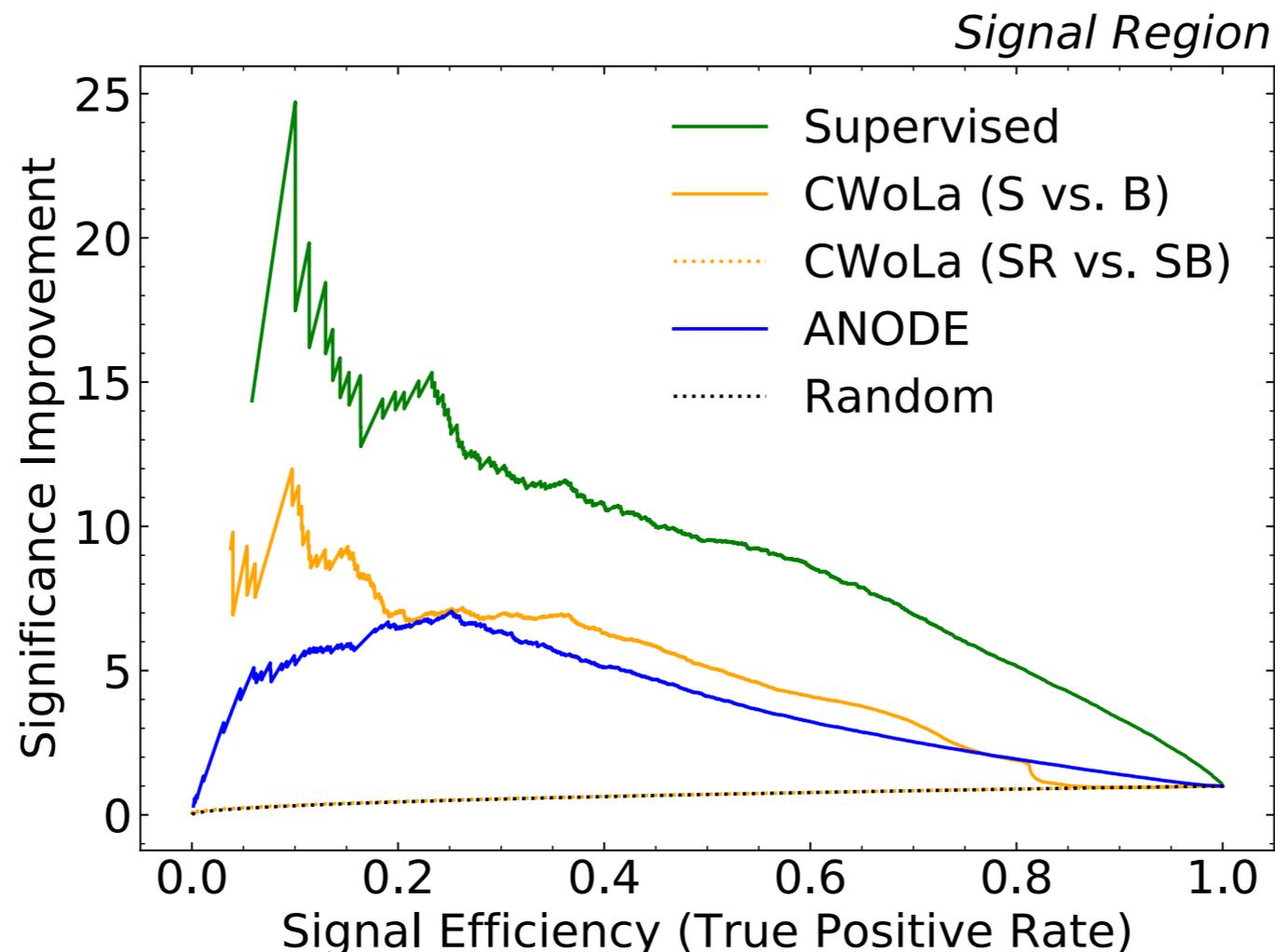
Ben Nachman & DS 2001.04990



The method works! ANODE is sensitive to the signal!

ANODE: Results on LHCO R&D Dataset

Ben Nachman & DS 2001.04990



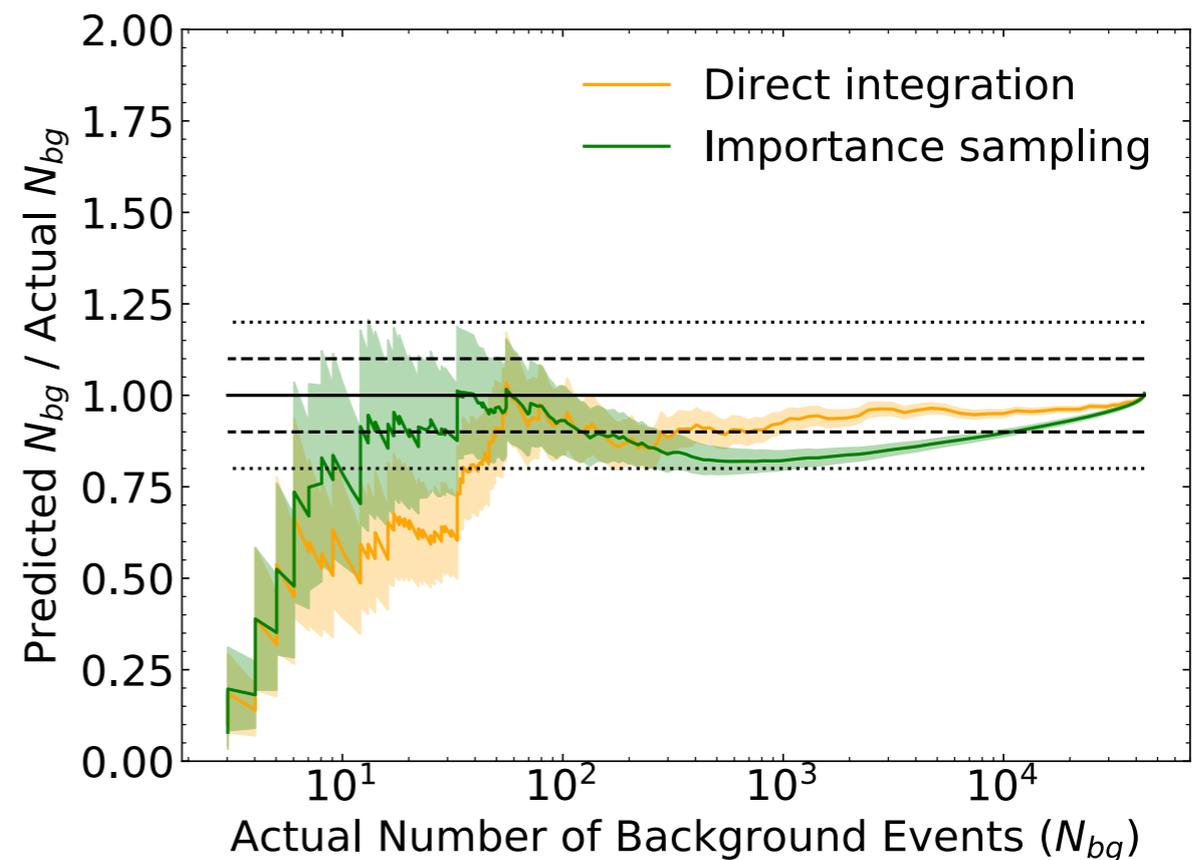
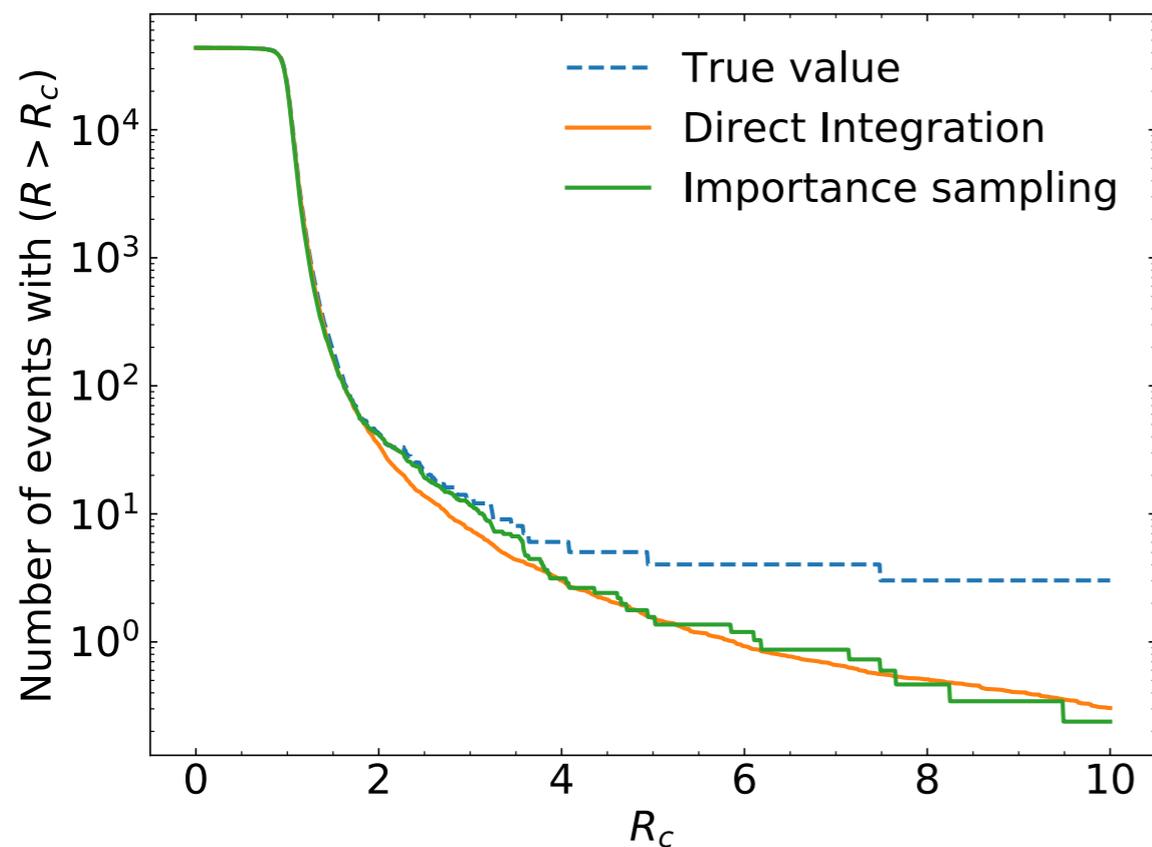
Can enhance the significance of the bump hunt by a factor of up to 7!

(For this feature set, the CWoLa independence assumptions are satisfied, and it outperforms ANODE. Shows the power of likelihood free methods.)

ANODE: Results on LHCO R&D Dataset

Ben Nachman & DS 2001.04990

Novel aspect of ANODE: can estimate backgrounds directly with $L(x|B_{\text{data}}; m \in \text{SR})$

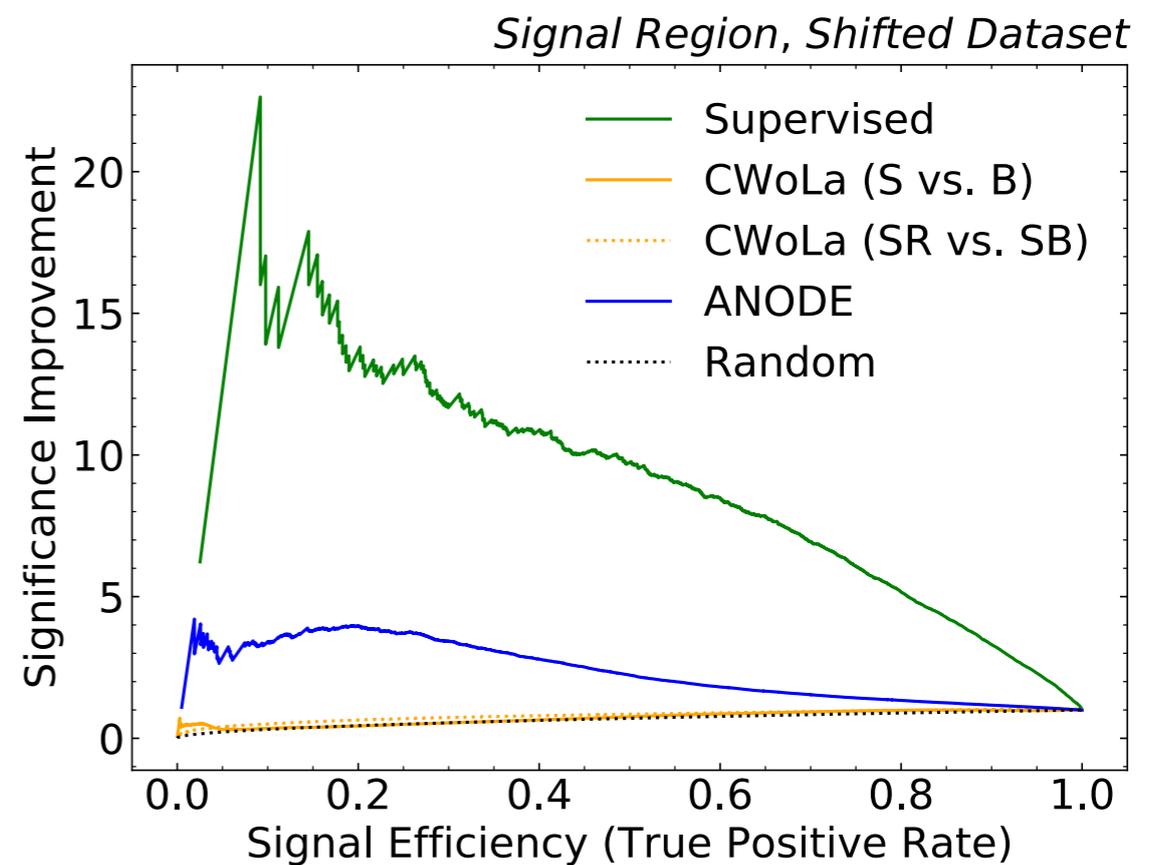
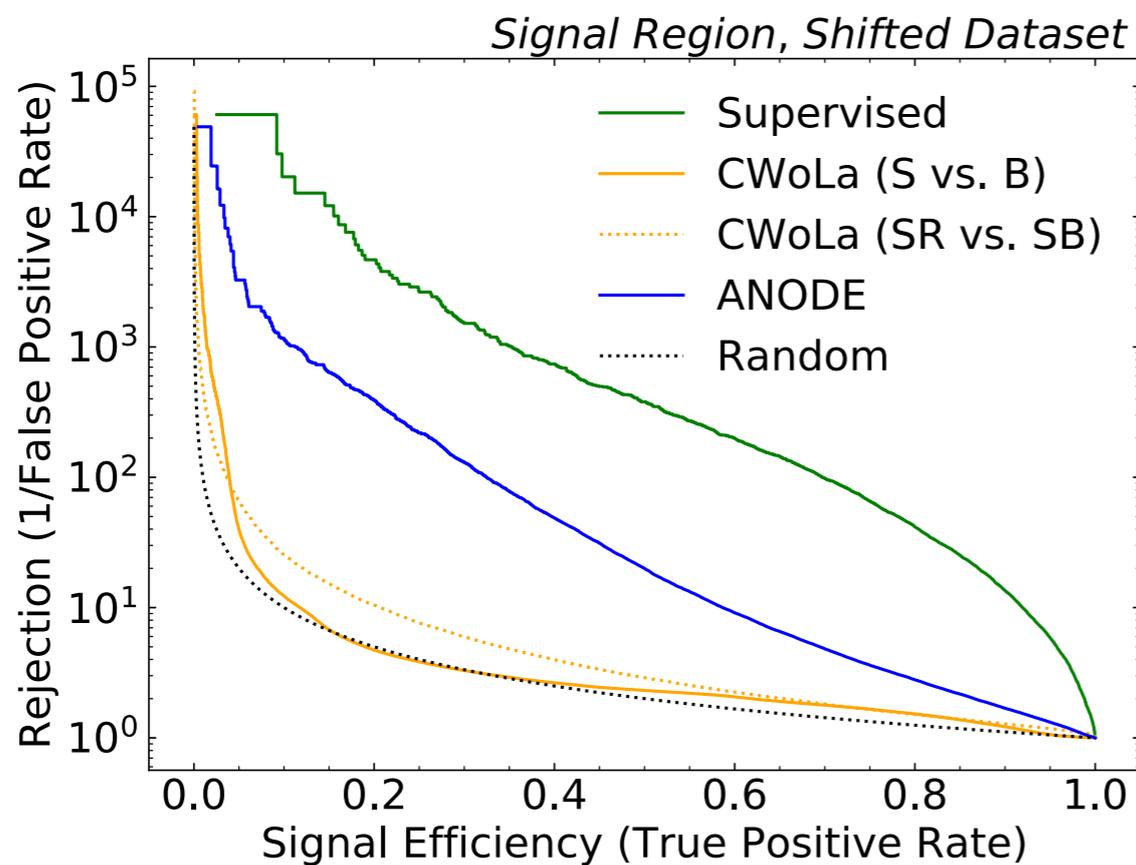


ANODE: Results on LHCO R&D Dataset

Ben Nachman & DS 2001.04990

Can also consider performance on a feature set which is not independent of m . We introduced artificial correlations just as proof of concept:

$$m_{J_{1,2}} \rightarrow m_{J_{1,2}} + c m_{JJ}$$



ANODE is robust while CWoLa completely fails!

LHC Olympics 2020: Submission format

A p-value associated with the dataset having no new particles (null hypothesis).

Short answer text
.....

As complete a description of the new physics as possible. For example: the masses and decay modes of all new particles (and uncertainties on those parameters).

Short answer text
.....

How many signal events (+uncertainty) are in the dataset (before any selection criteria).

Short answer text
.....

Please consider submitting plots or a Jupyter notebook! (these will be private and used only for the presentation / documentation at the end)

 Add file

Overview of submissions

- 10 groups submitted results on box 1
- 4 of these groups also submitted results on boxes 2 & 3
- A number of additional groups could not finish the challenge in time but got results on the R&D dataset
- 7 of these groups gave talks about their methods and results at the ML4Jets2020 conference

Overview of submissions

People tried both supervised and unsupervised methods.

Methods used included

- Autoencoders
- CWoLa hunting
- PCA outlier detection
- LSTM
- CNN+BDT
- variational RNNs for anti-QCD tagging
- density estimation
- biological neural network
- ...

Box 1

Signal: 834 events

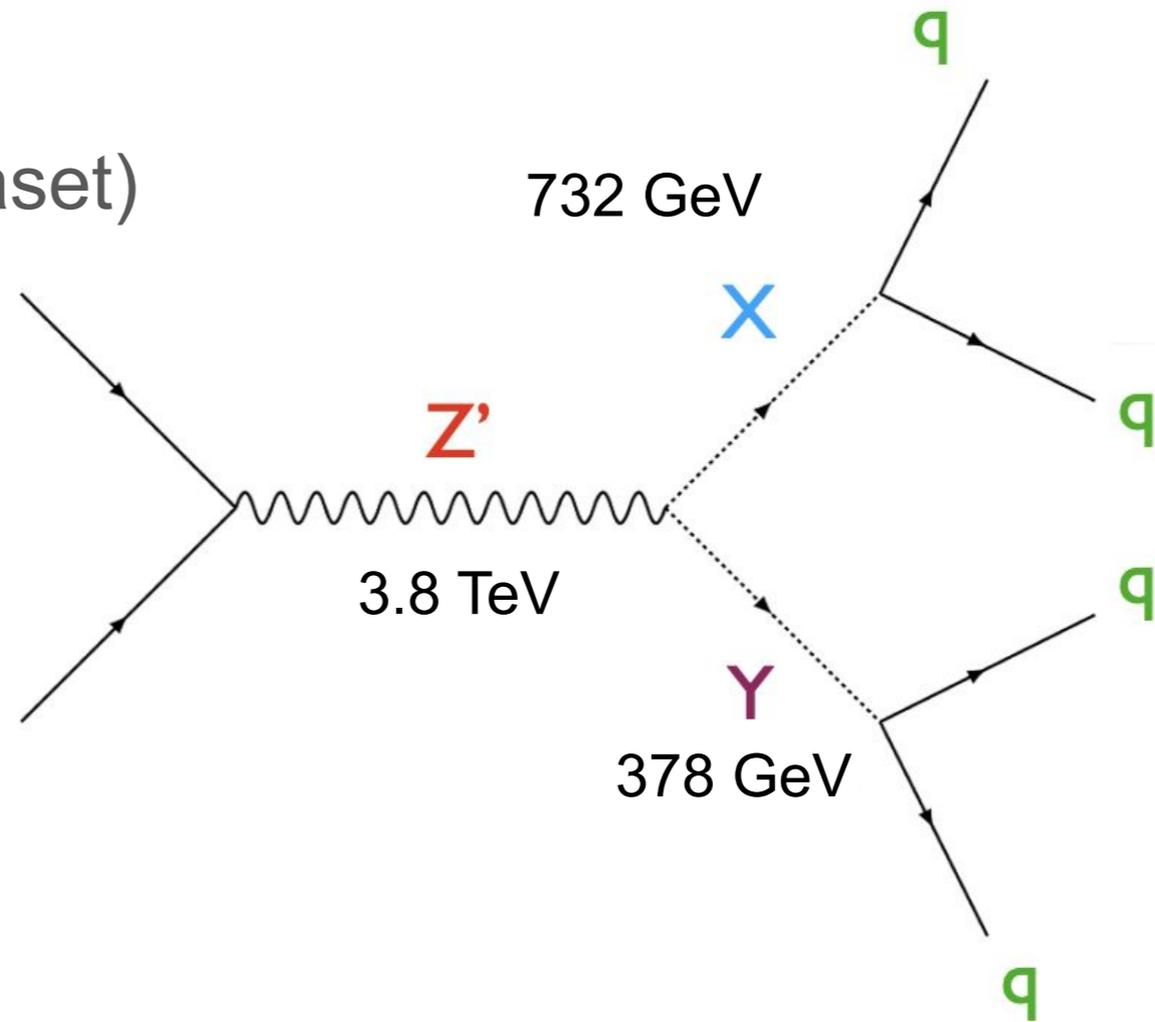
$Z' \rightarrow XY; X, Y \rightarrow qq$

(same topology as R&D dataset)

$m_{Z'} = 3823 \text{ GeV}$

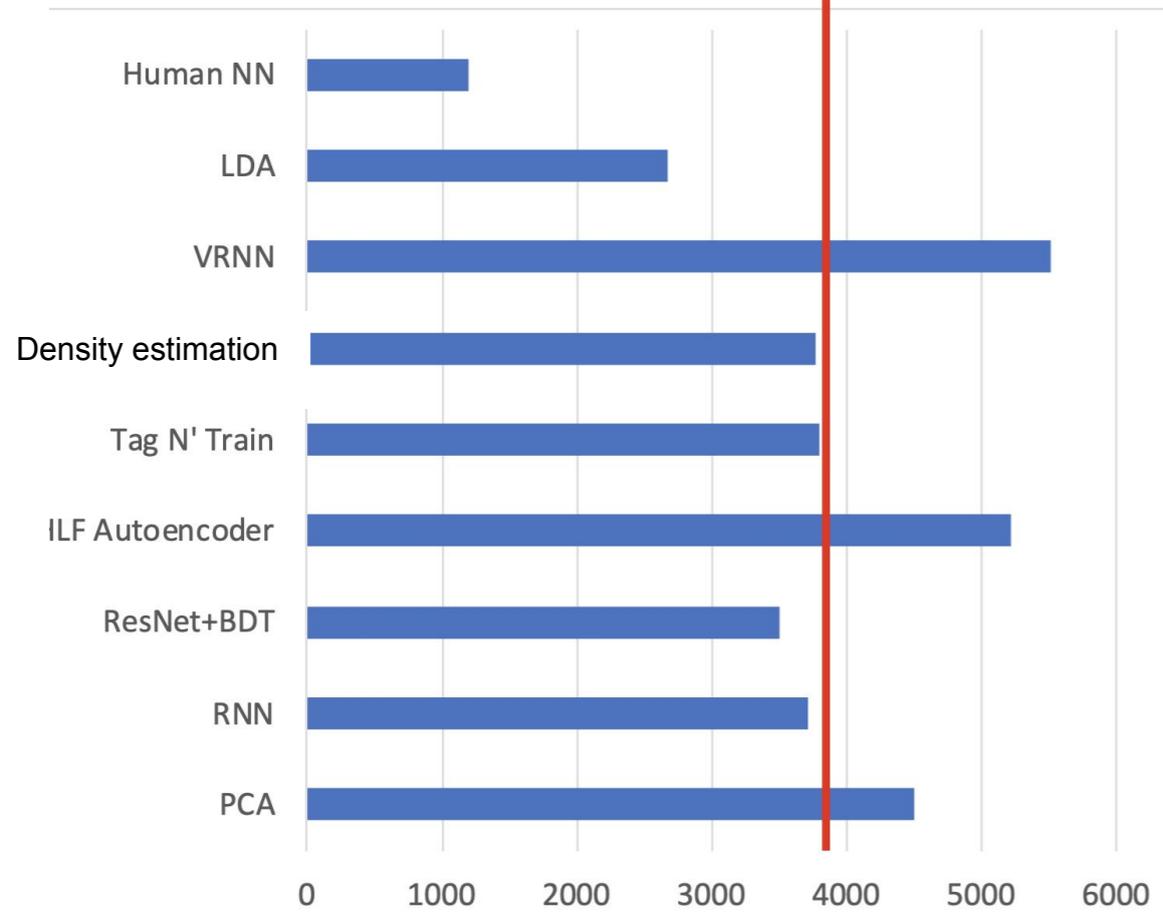
$m_X = 732 \text{ GeV}$

$m_Y = 378 \text{ GeV}$

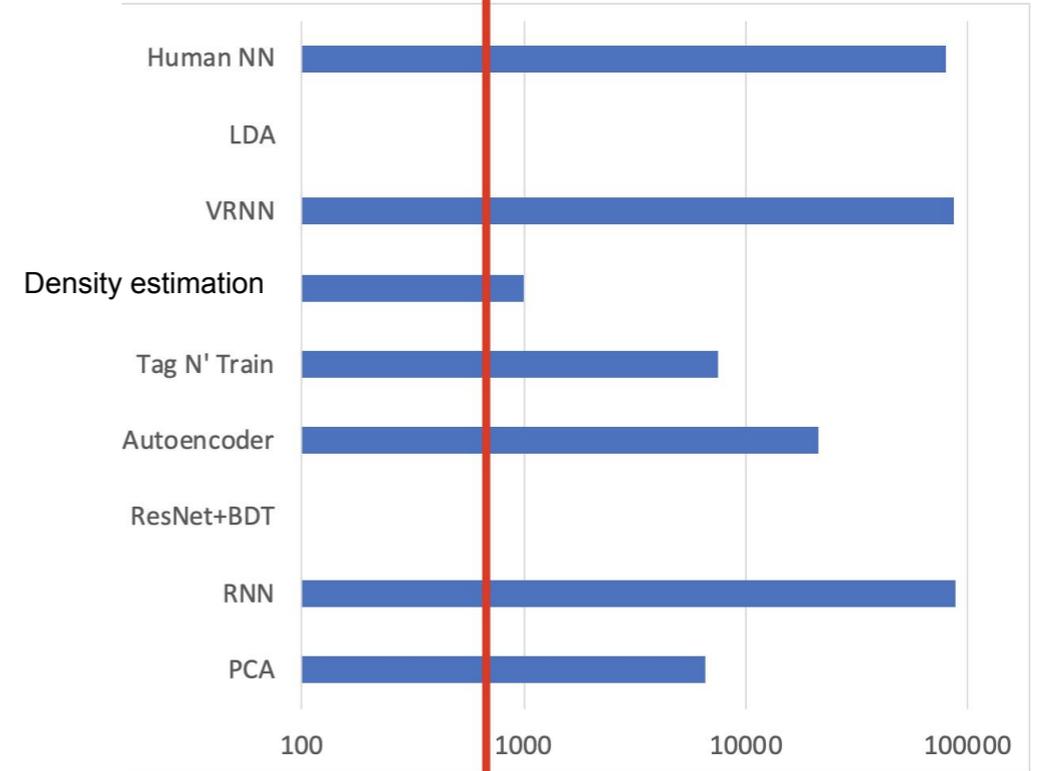


We revealed the answer at the ML4Jets2020 conference
in early January

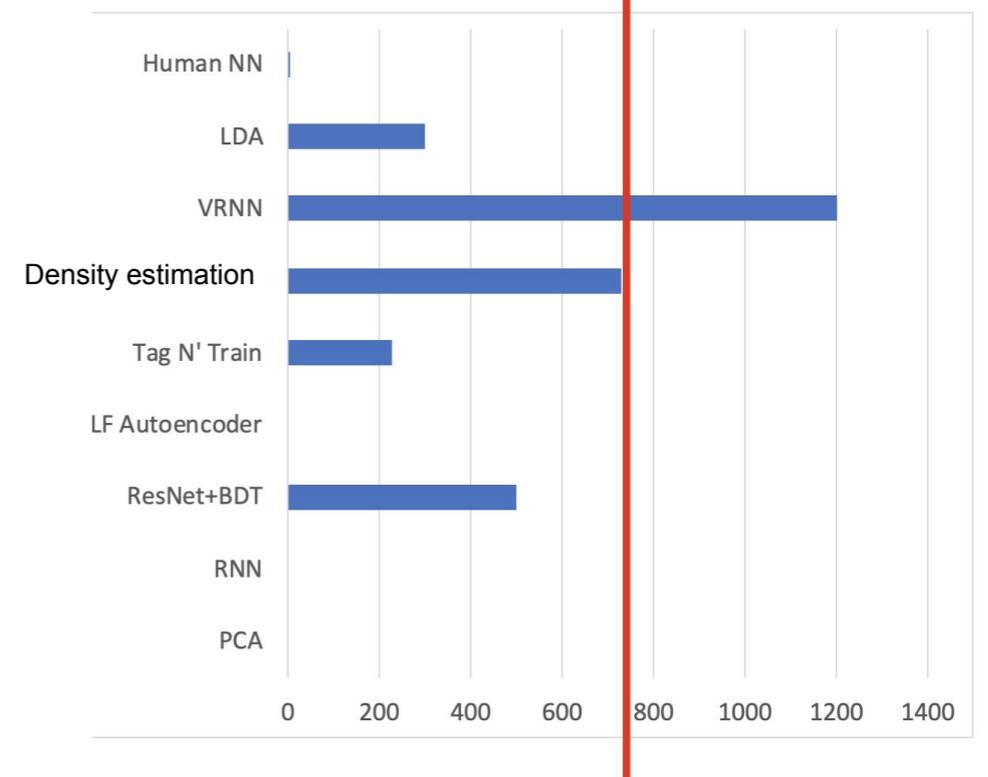
M = 3823 GeV

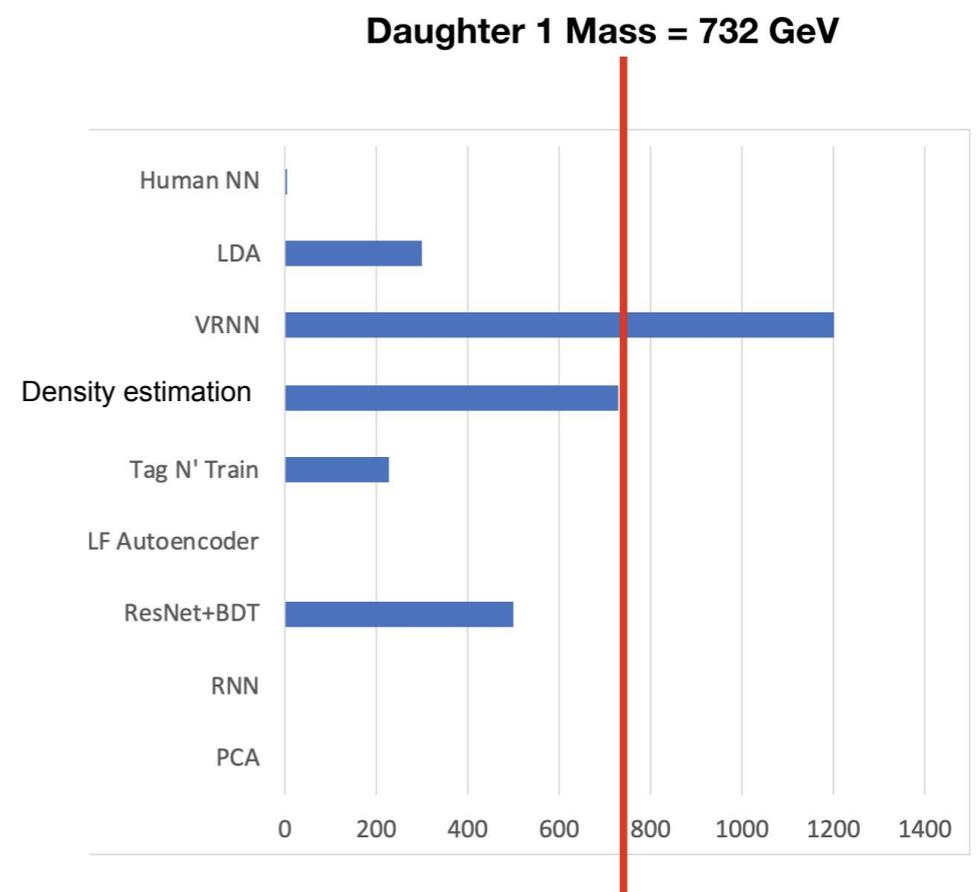
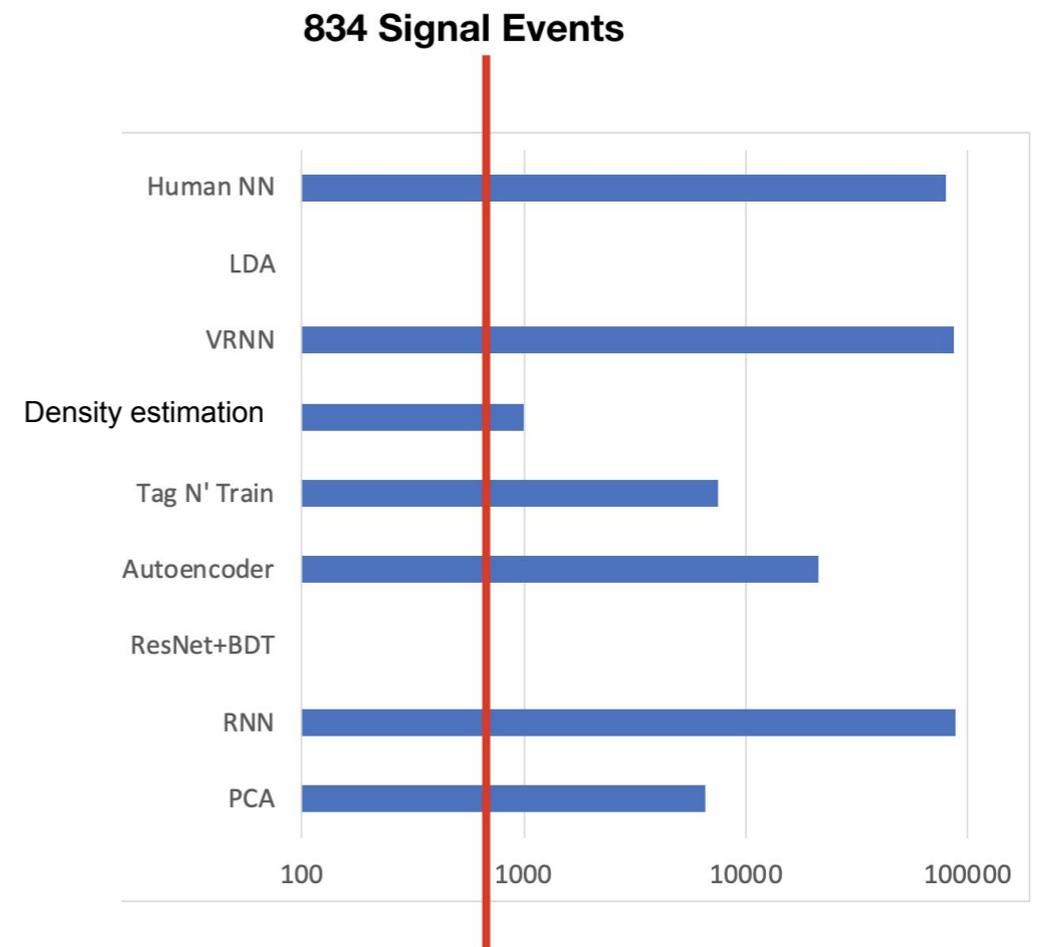
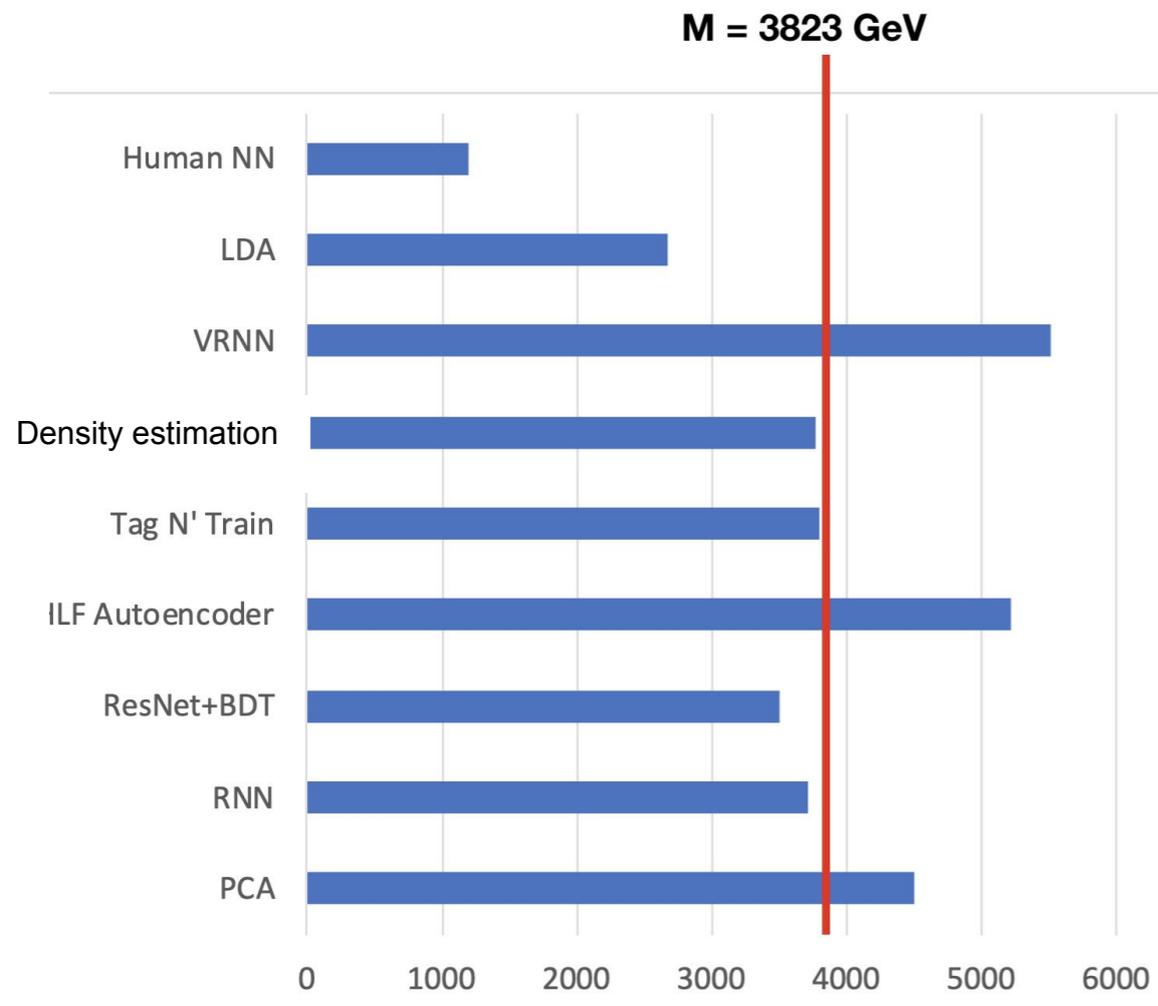


834 Signal Events



Daughter 1 Mass = 732 GeV





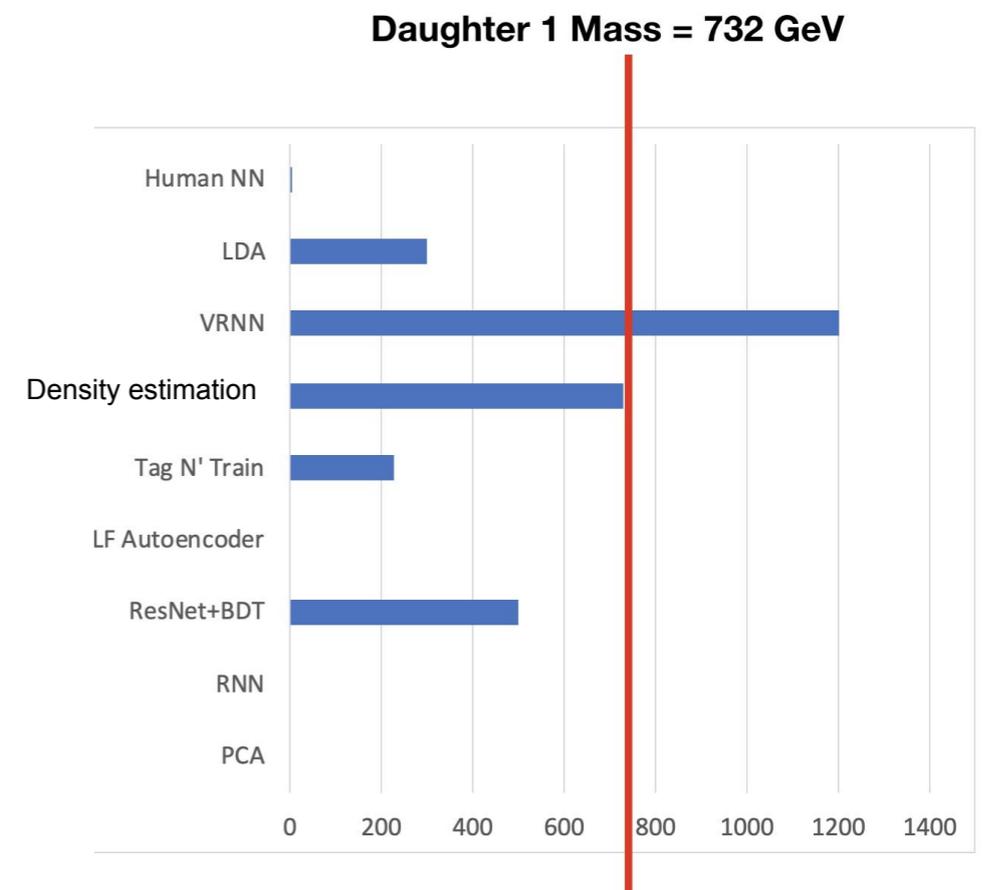
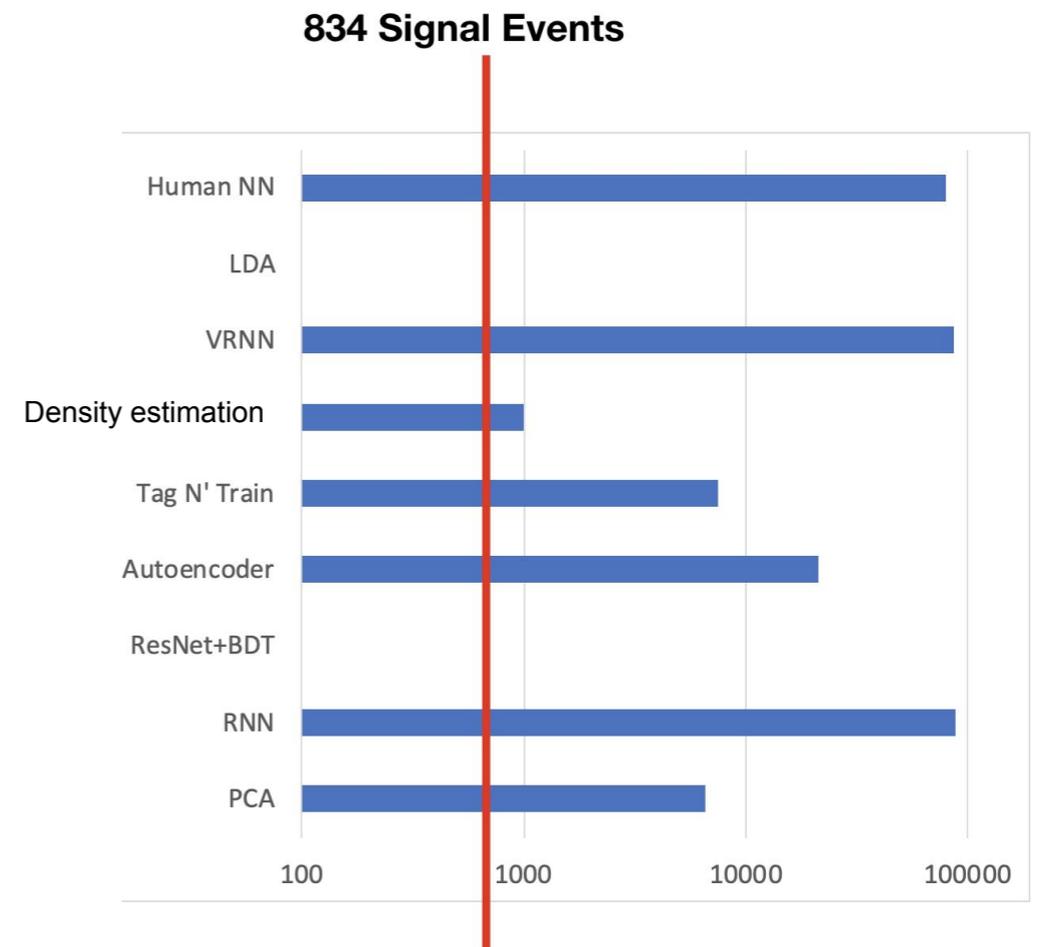
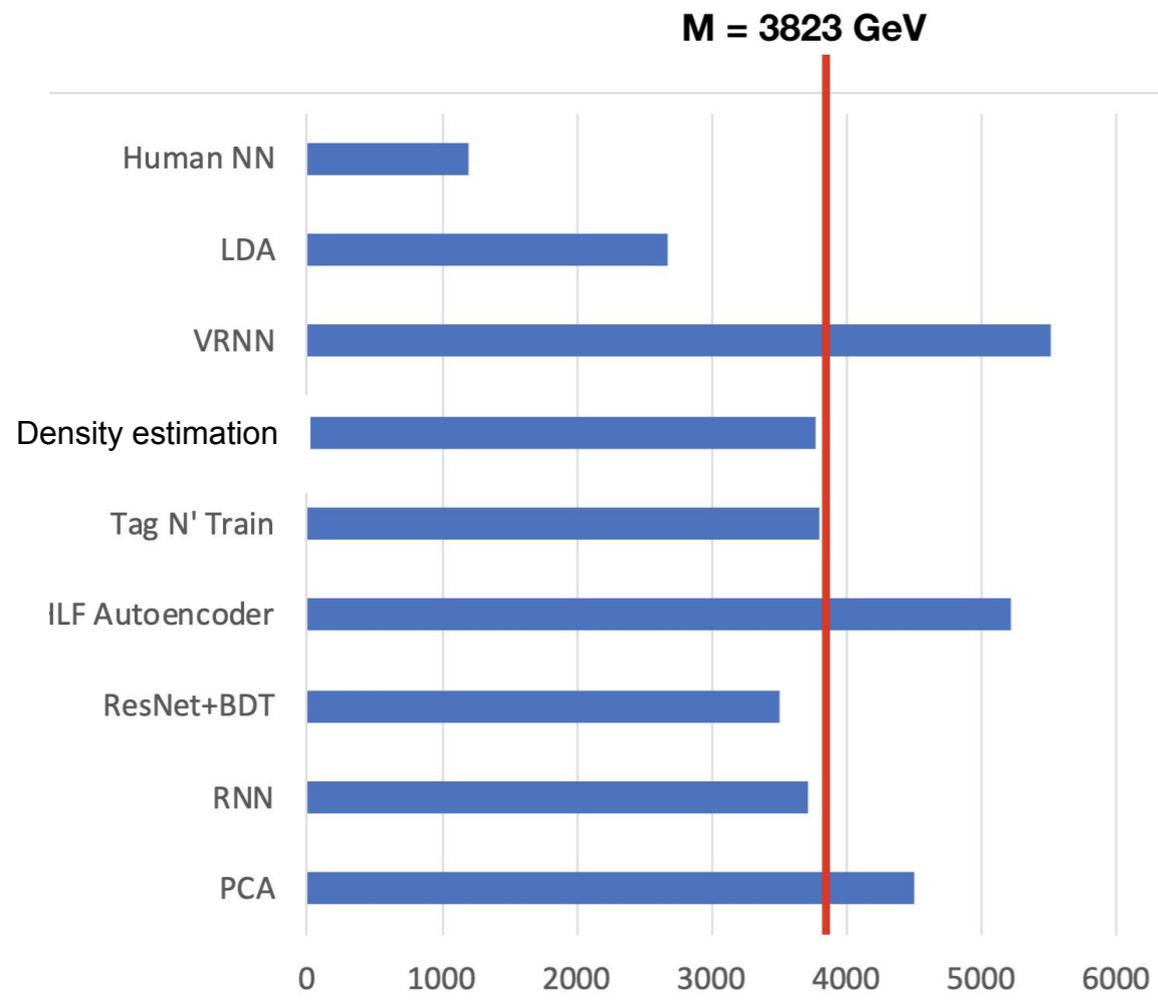
A clear winner emerged:

Conditional density estimation for anomaly detection

George Stein, Uros Seljak, Biwei Dai, He Jia



Used the ANODE method with a novel density estimator!



In second place:

Tag N' Train

Oz Amram & Cristina Mantilla Suarez (Johns Hopkins)



Used a combination of autoencoders and CWoLa hunting

LHCO2020: Summer Games

<https://indico.desy.de/indico/event/25341/>

Stay tuned for more on the LHCO 2020...

We will be organizing a 1-day mini-workshop on anomaly detection in Hamburg the Saturday before BOOST (July 18).

There the answers for Boxes 2 and 3 will be revealed.

We will also discuss plans for a community paper on new methods for anomaly detection and the LHCO2020.

Please come and join us!

Conclusions

These are exciting times for anomaly detection in HEP.

Many new approaches making use of unsupervised ML are being developed by theorists and experimentalists.

Model independent searches have a bright future at the LHC. Maybe this is how we will finally discover the new physics!

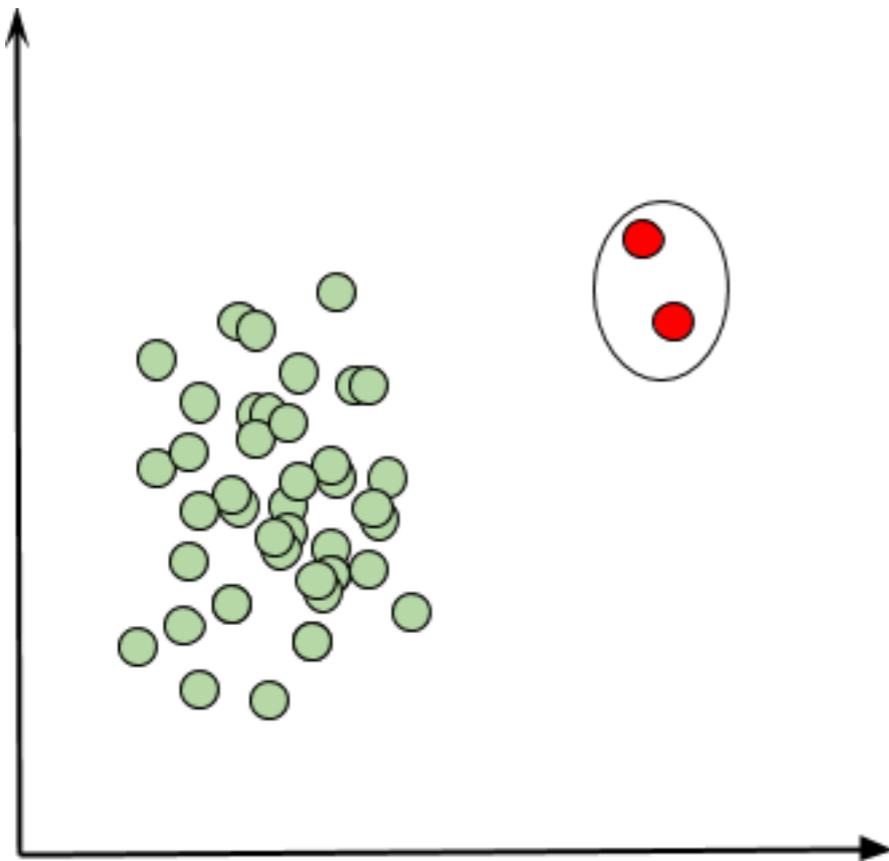
These methods also have potential applications beyond HEP. For example, ANODE is a completely general method for finding localized overdensities in high dimensional datasets. One can imagine many uses for such a method!

Thanks for your attention!

Anomaly Detection at the LHC

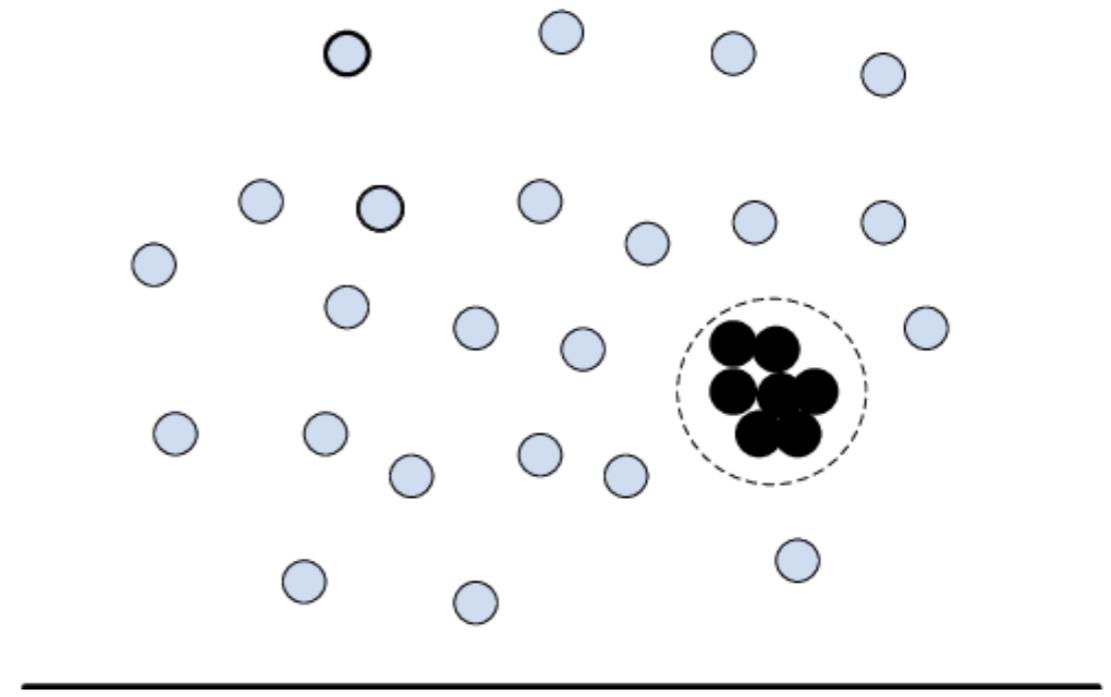
In the broader ML field, there are two types of anomaly detection

point outlier, out of sample anomaly,
“zero-background search”



easy

collective outlier, population anomaly
“bump hunt”



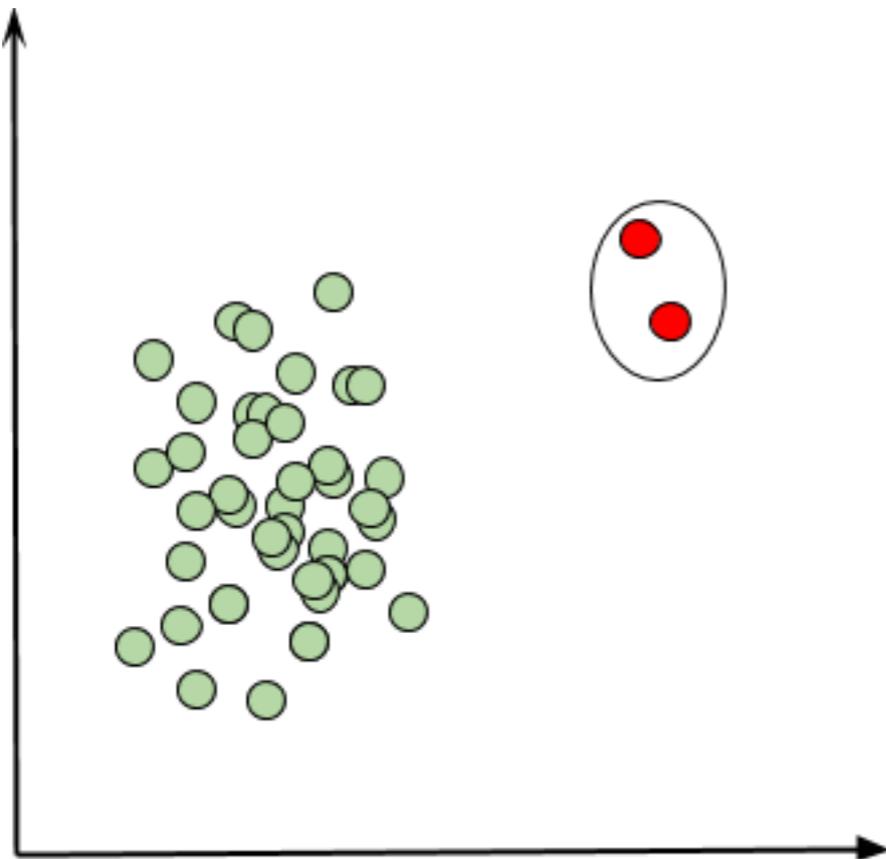
hard

Anomaly Detection at the LHC



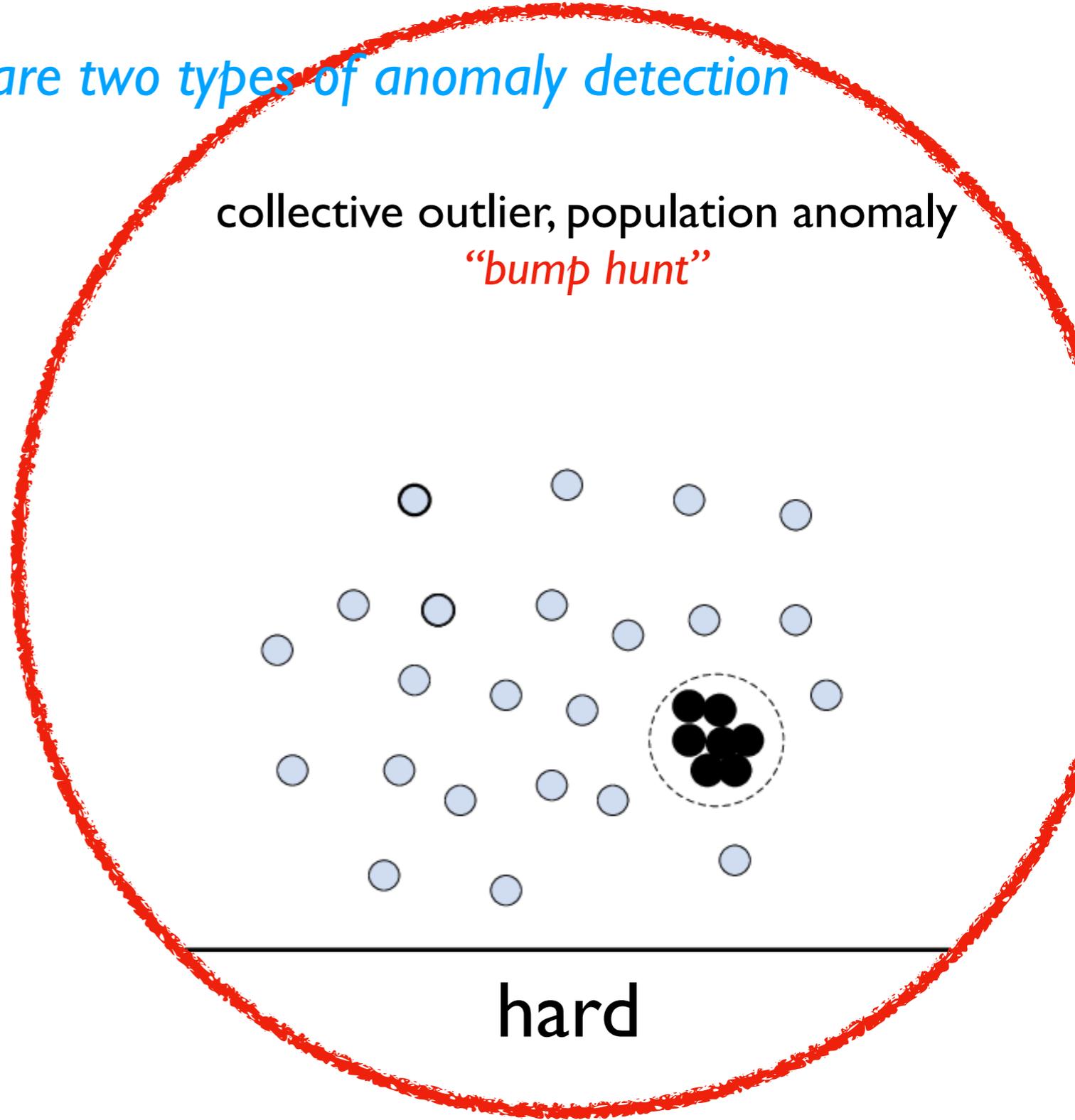
In the broader ML field, there are two types of anomaly detection

point outlier, out of sample anomaly,
“zero-background search”



easy

collective outlier, population anomaly
“bump hunt”



hard

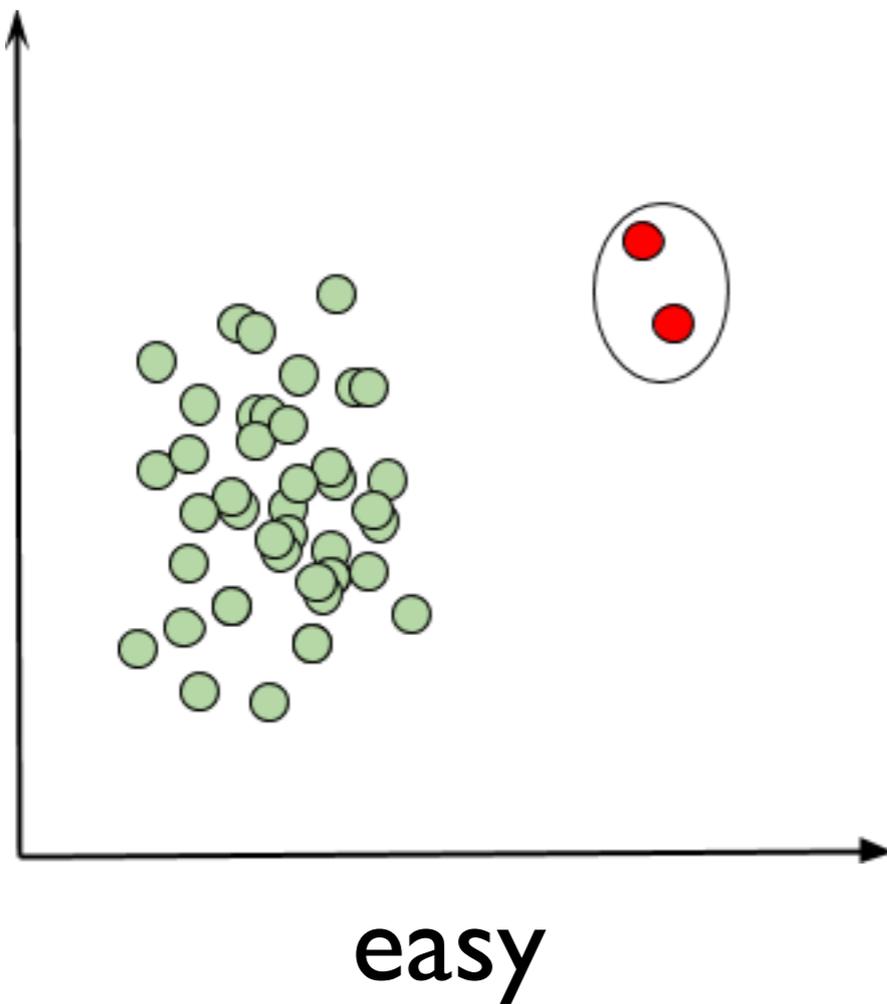
Anomaly Detection at the LHC



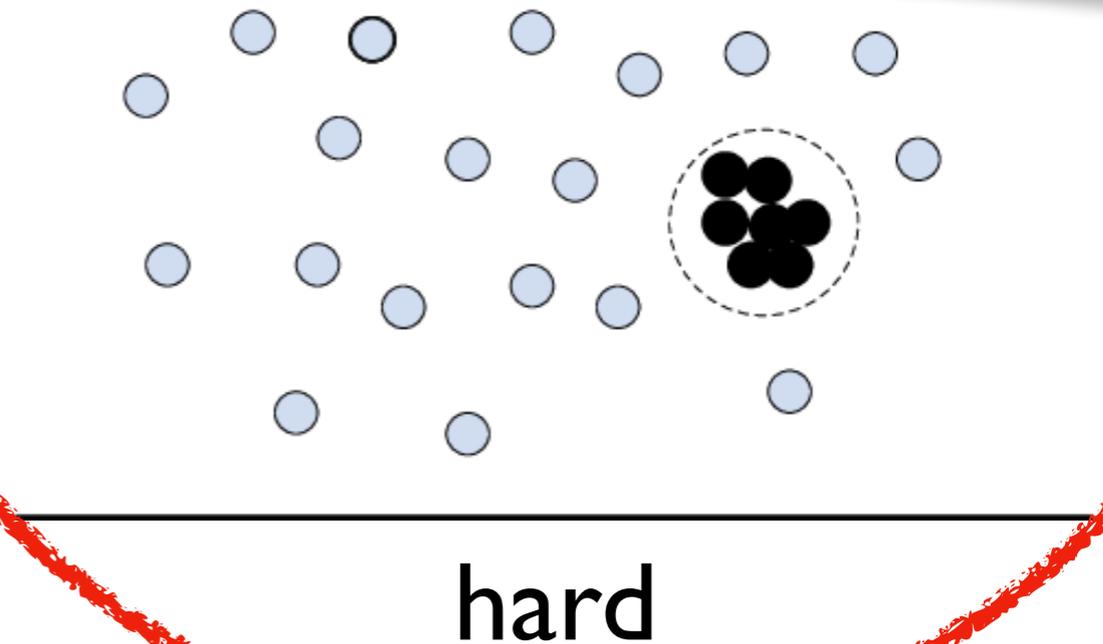
In the broader ML field, there are two types of anomaly detection

point outlier, out of sample anomaly,
“zero-background search”

collective outlier, population anomaly
“bump hunt”

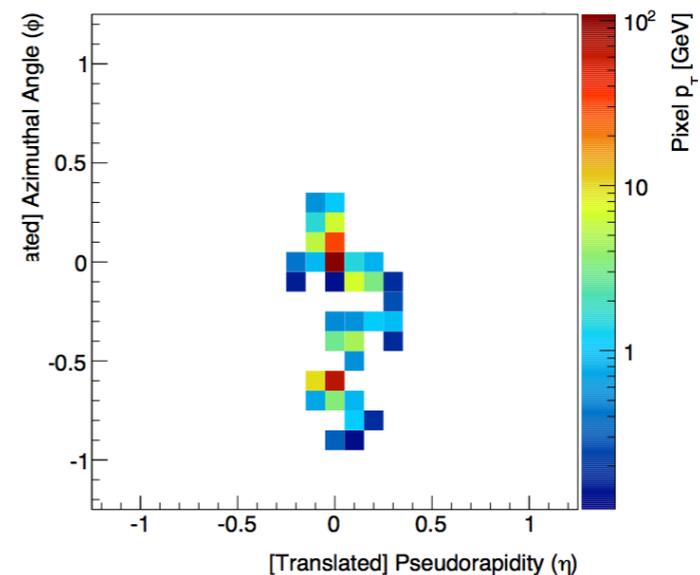
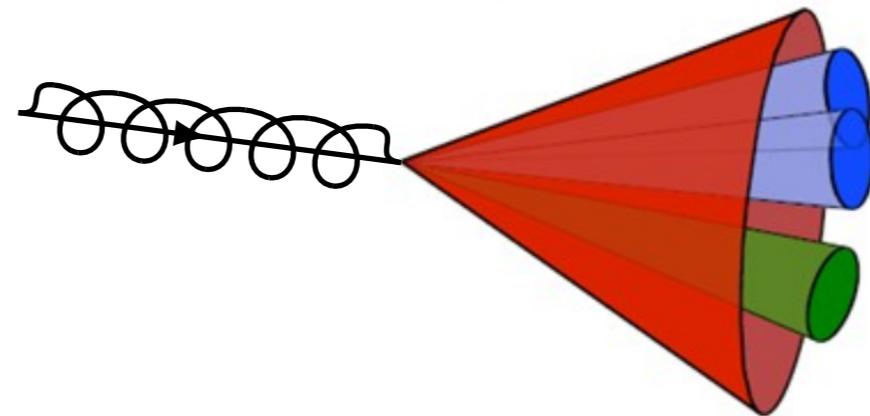
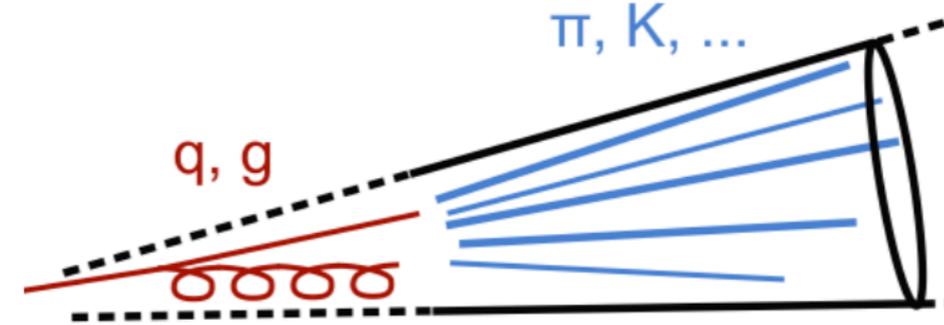


Background estimation is a key component of anomaly detection in HEP



Sample definitions

- Background: QCD jets
(p_T : 800-900 GeV, $|\eta| < 1$, anti-kt $R=1$)
- Signals:
 - All-hadronic tops
 - 400 GeV gluinos decaying via RPV
- We formed jet images in η and ϕ with a pixel resolution and intensity given by the calorimeter towers.



Searching for NP with deep autoencoders

Heimel et al I808.08979; Farina, Nakai & DS I808.08992

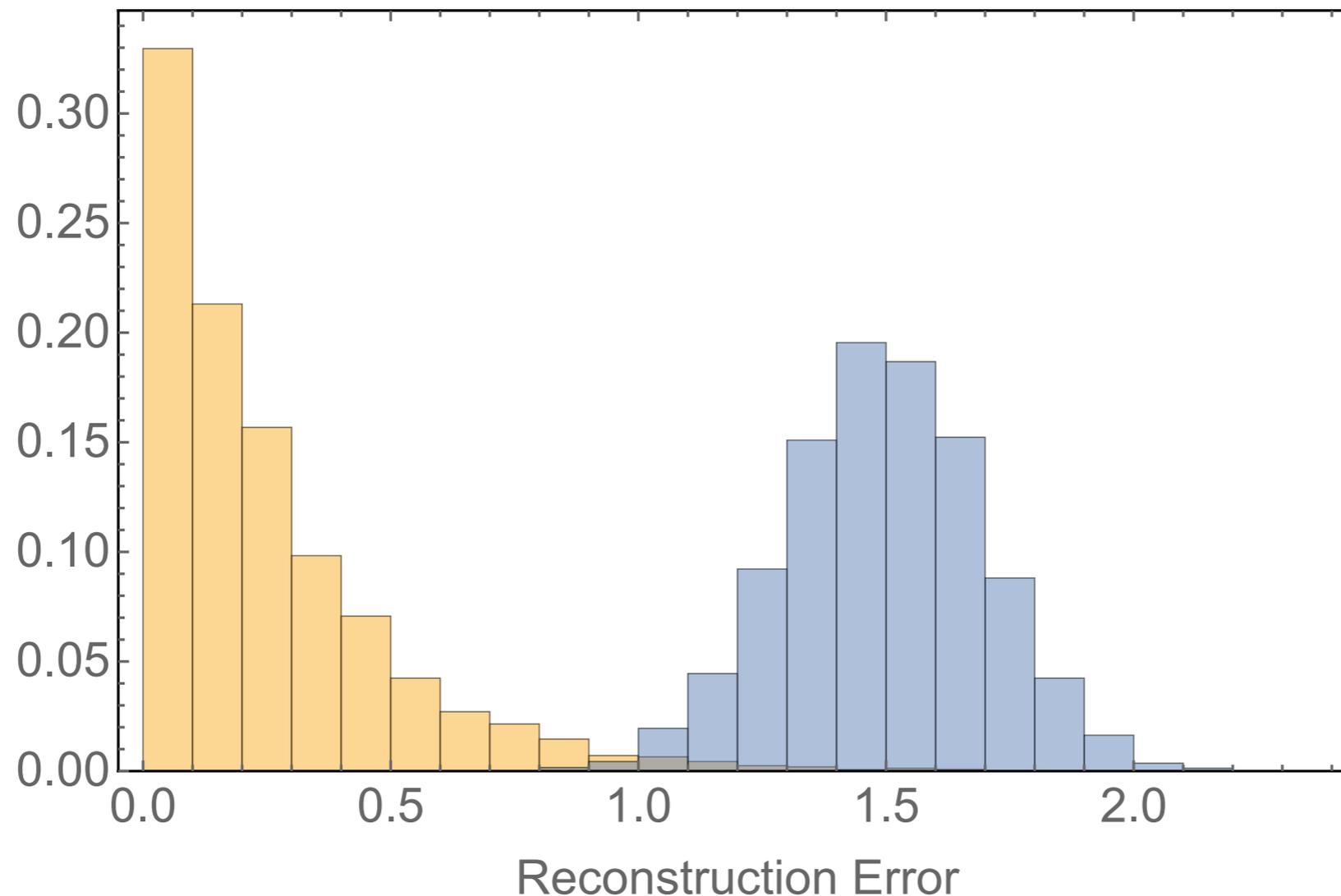
Loss function for autoencoder:
“reconstruction error”

$$L = \frac{1}{N} \sum_{i=1}^N (x_i^{in} - x_i^{out})^2$$

Searching for NP with deep autoencoders

Heimel et al I808.08979; Farina, Nakai & DS I808.08992

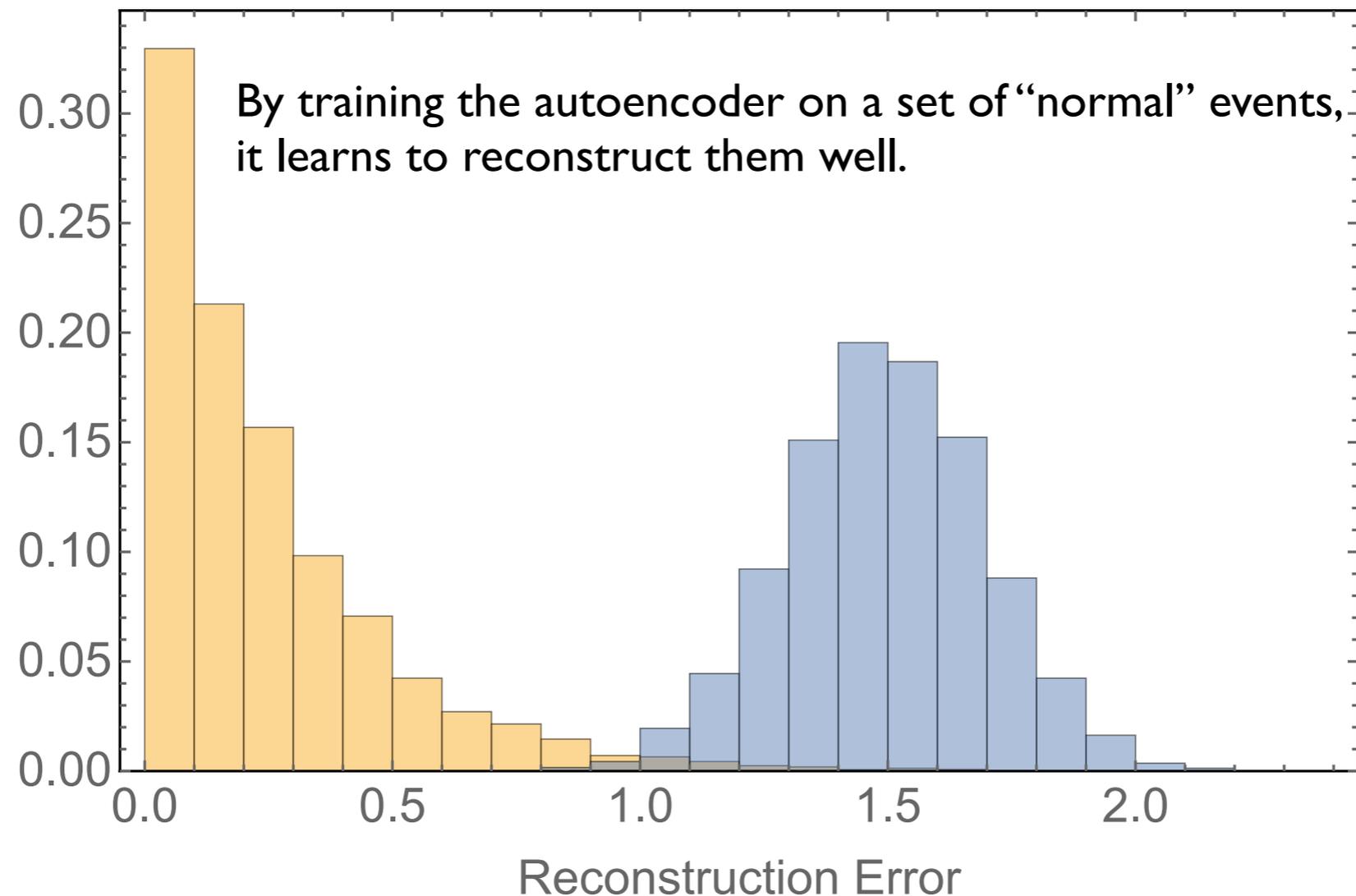
Loss function for autoencoder: $L = \frac{1}{N} \sum_{i=1}^N (x_i^{in} - x_i^{out})^2$
“reconstruction error”



Searching for NP with deep autoencoders

Heimel et al I808.08979; Farina, Nakai & DS I808.08992

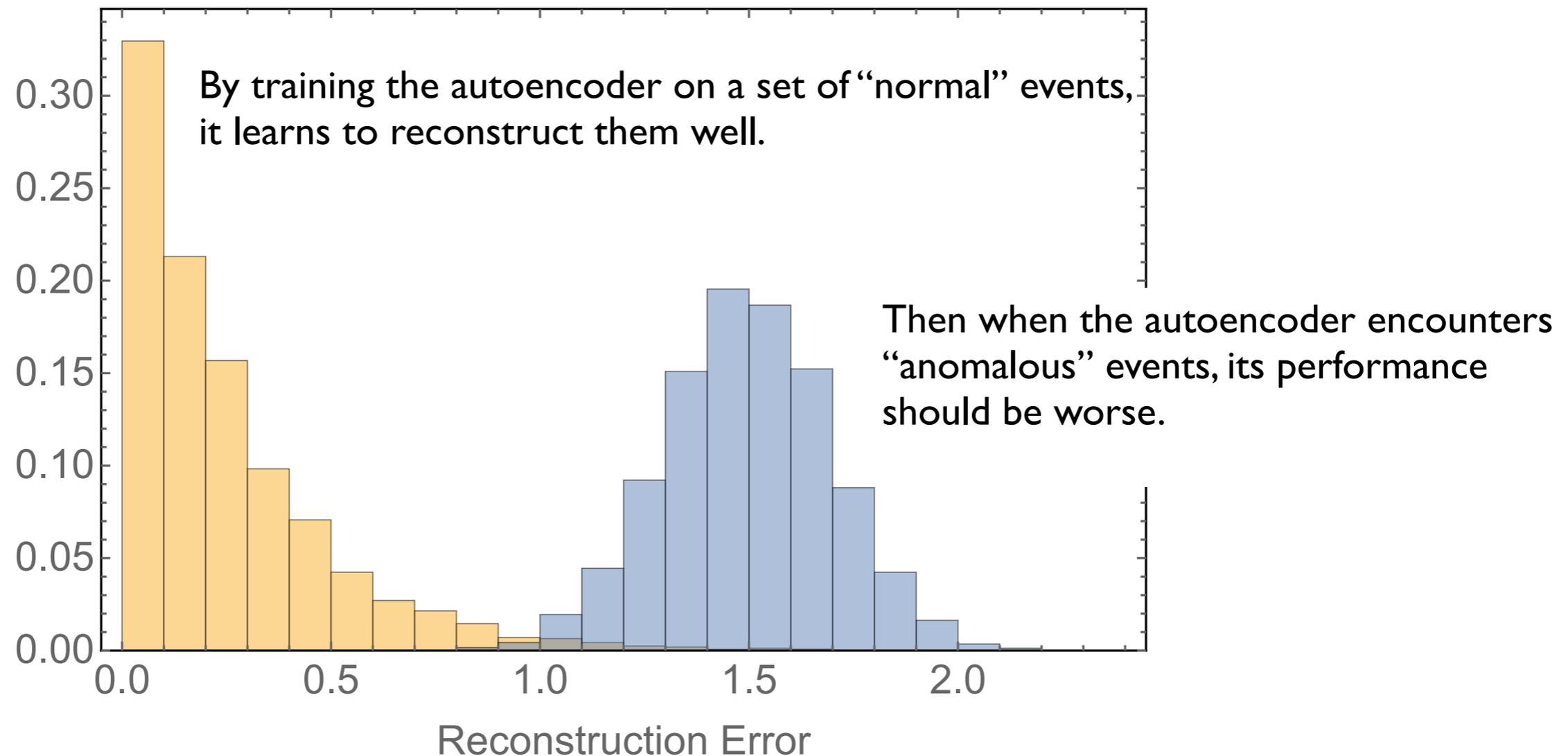
Loss function for autoencoder: $L = \frac{1}{N} \sum_{i=1}^N (x_i^{in} - x_i^{out})^2$
“reconstruction error”



Searching for NP with deep autoencoders

Heimel et al I808.08979; Farina, Nakai & DS I808.08992

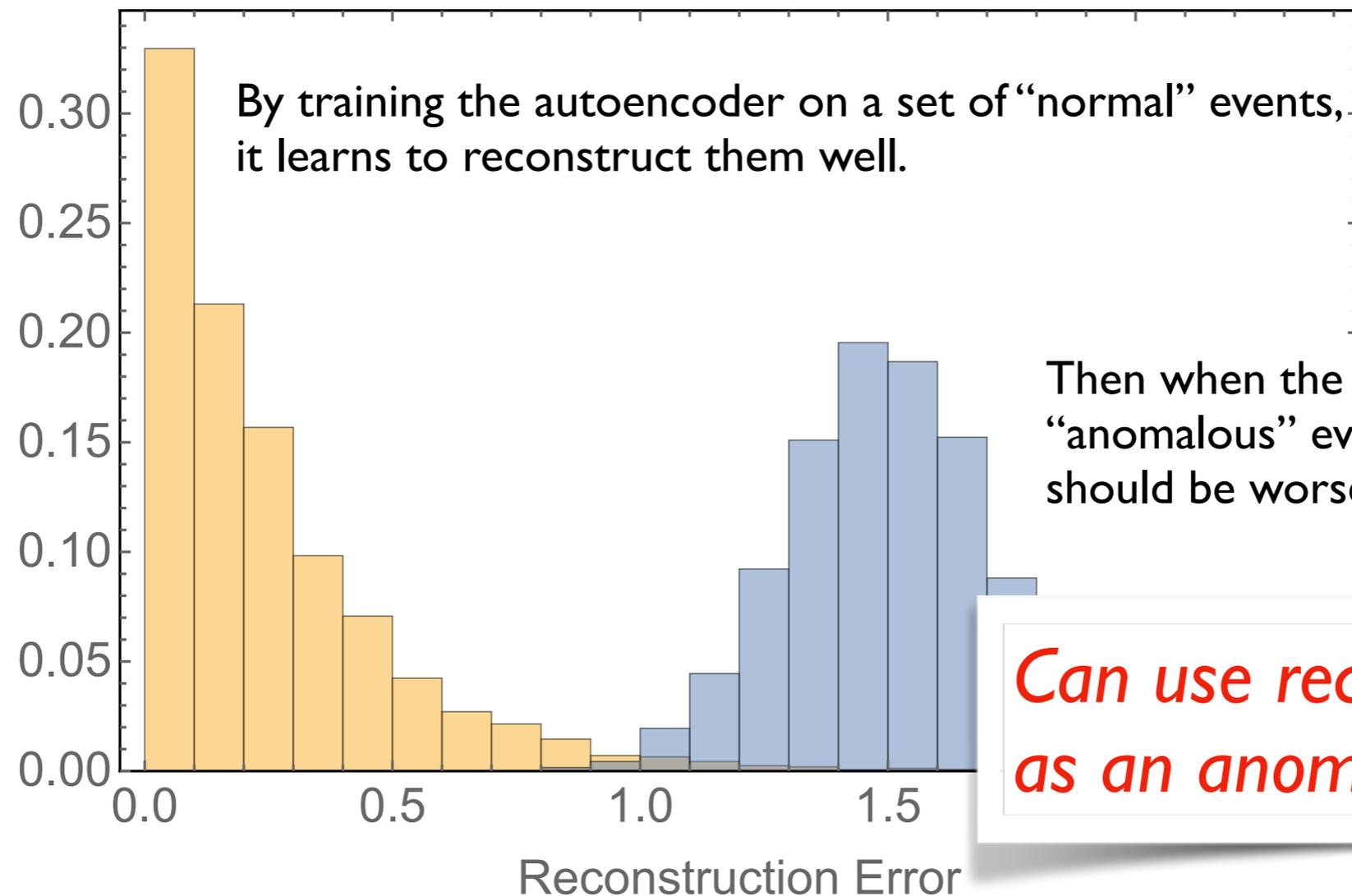
Loss function for autoencoder: $L = \frac{1}{N} \sum_{i=1}^N (x_i^{in} - x_i^{out})^2$
“reconstruction error”



Searching for NP with deep autoencoders

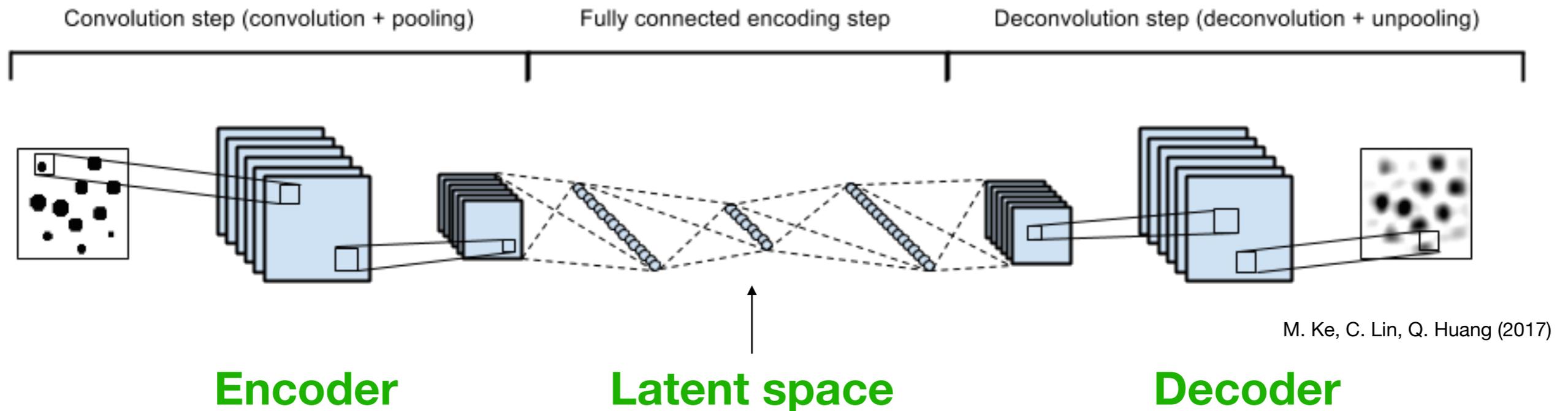
Heimel et al I808.08979; Farina, Nakai & DS I808.08992

Loss function for autoencoder: $L = \frac{1}{N} \sum_{i=1}^N (x_i^{in} - x_i^{out})^2$
“reconstruction error”



Can use reconstruction error as an anomaly threshold!

Autoencoder architecture



I28C3-MP2-I28C3-MP2-I28C3-32N-6N-32N-I2800N-I28C3-US2-I28C3-US2-I C3

Our primary autoencoder used **convolutional neural networks (CNNs)** for encoding and decoding the jet images.

We also considered autoencoders based on PCA and simple DNNs.

Many more architectures are possible.

Choosing the latent dimension

Should choose the latent dimension in an unsupervised manner
(ie without optimizing on a specific signal)

- d too large \rightarrow autoencoder becomes identity transform
- d too small \rightarrow autoencoder cannot learn all the features

Choosing the latent dimension

Should choose the latent dimension in an unsupervised manner
(ie without optimizing on a specific signal)

- d too large \rightarrow autoencoder becomes identity transform
- d too small \rightarrow autoencoder cannot learn all the features

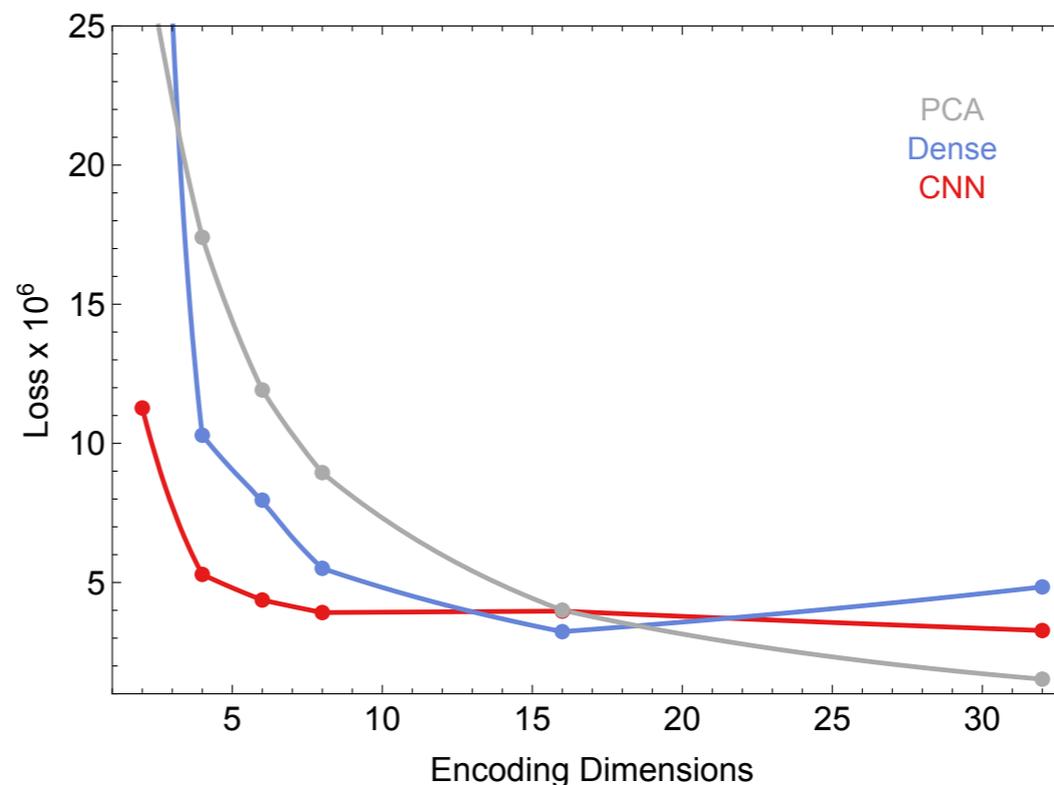
 *idea: look for saturation point as d is increased*

Choosing the latent dimension

Should choose the latent dimension in an unsupervised manner
(ie without optimizing on a specific signal)

- d too large \rightarrow autoencoder becomes identity transform
- d too small \rightarrow autoencoder cannot learn all the features

\rightarrow idea: look for saturation point as d is increased

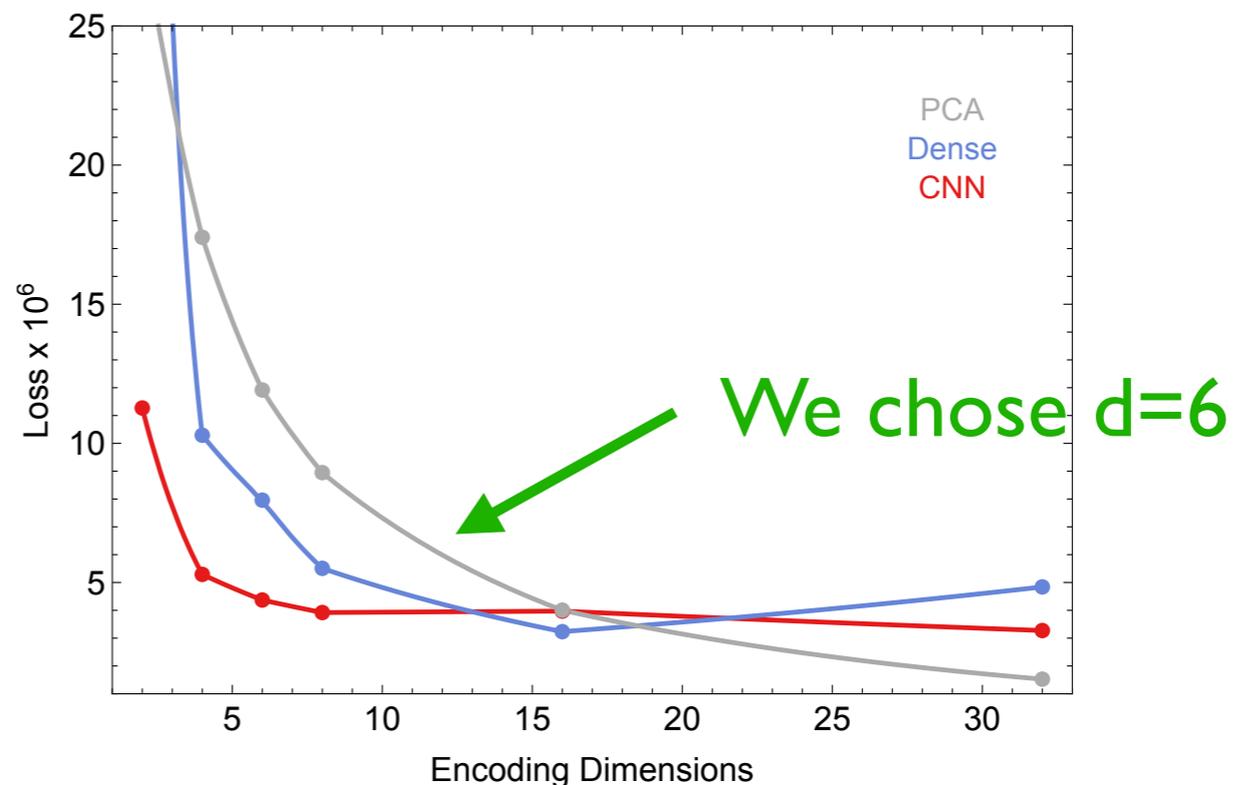


Choosing the latent dimension

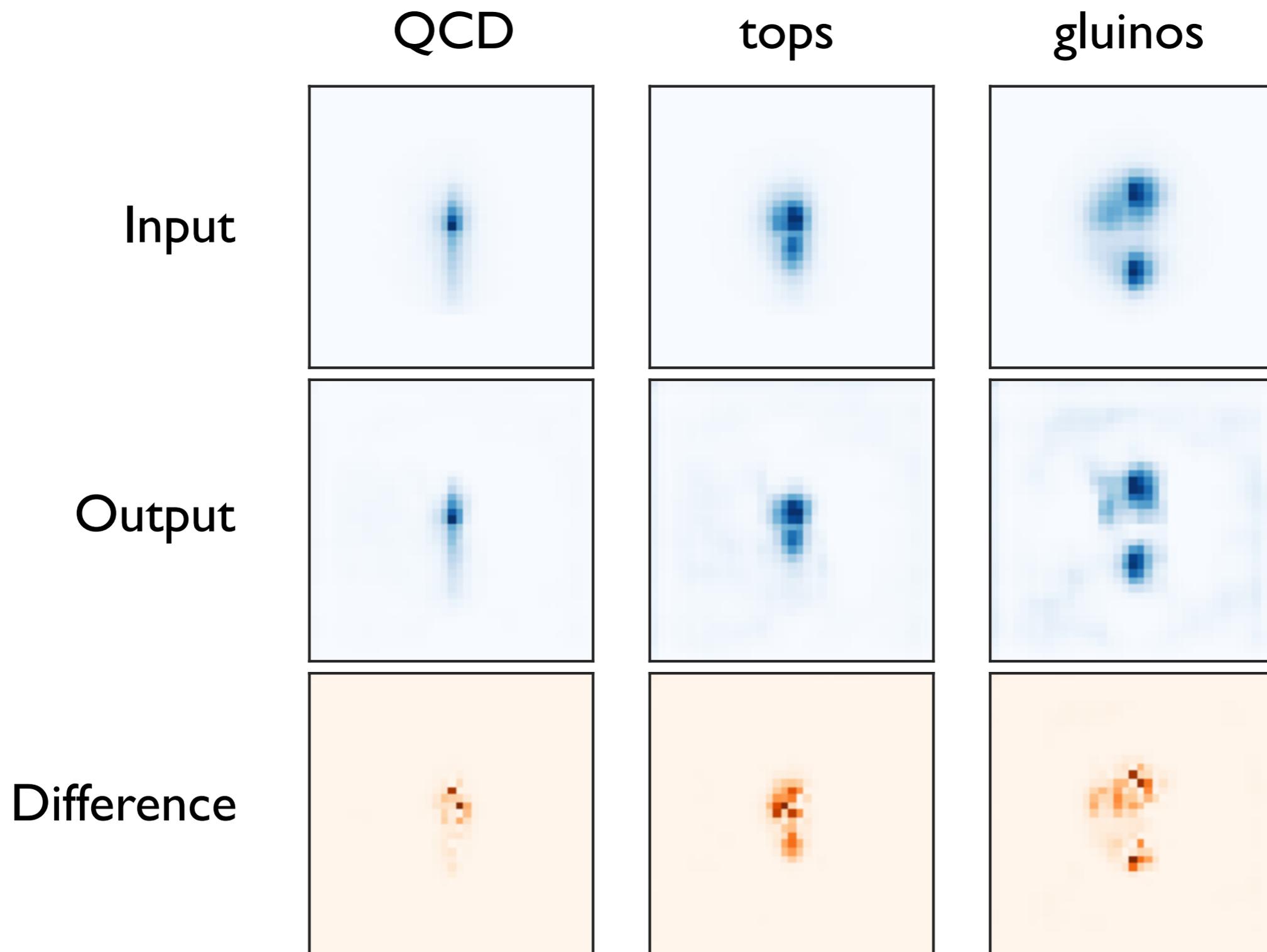
Should choose the latent dimension in an unsupervised manner
(ie without optimizing on a specific signal)

- d too large \rightarrow autoencoder becomes identity transform
- d too small \rightarrow autoencoder cannot learn all the features

➔ *idea: look for saturation point as d is increased*



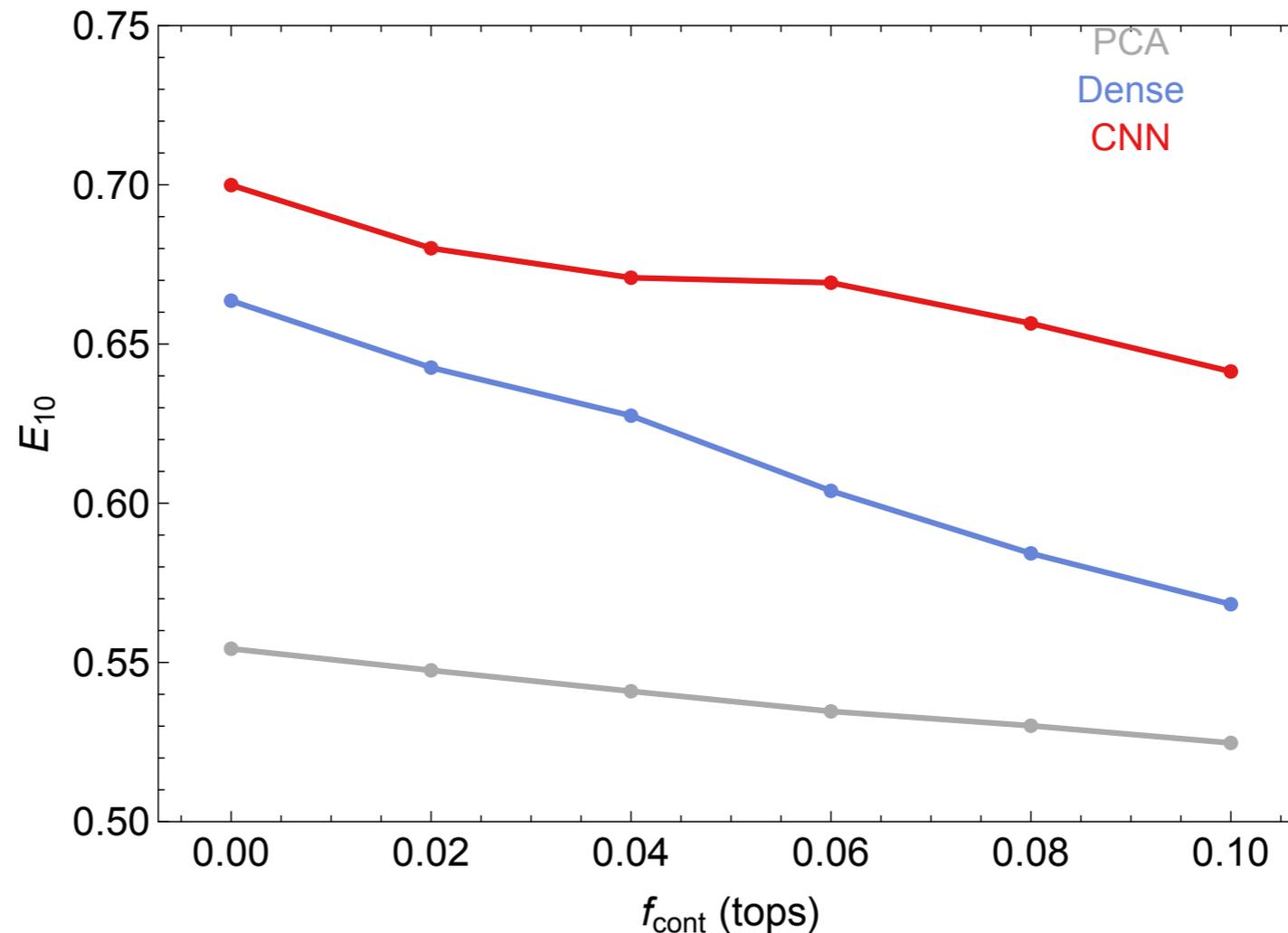
Performance: weakly supervised mode



Fully unsupervised mode

Can also train on QCD background “contaminated” with a small fraction of signal. **This could be representative of actual data.**

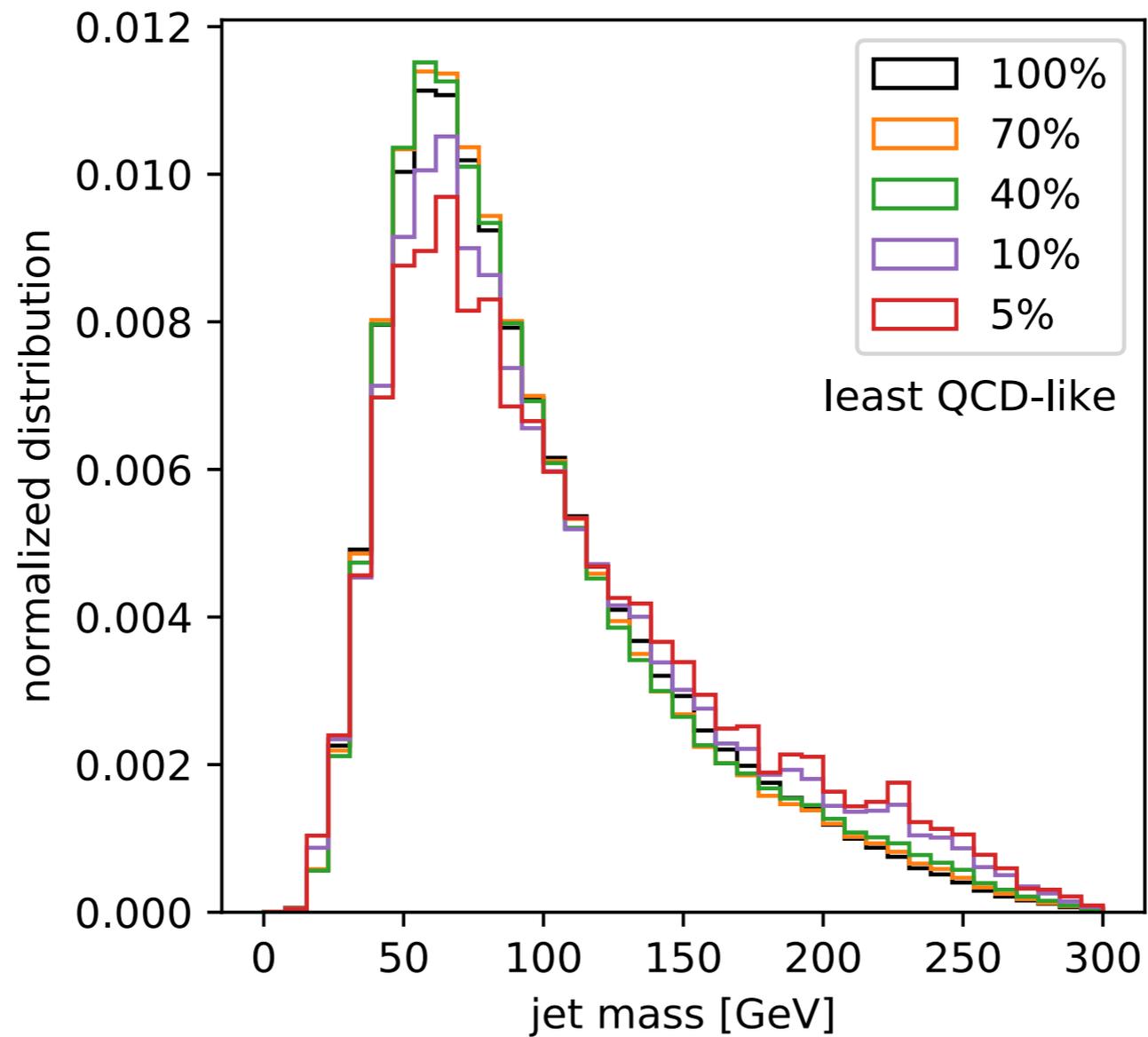
(E_{10} = signal efficiency
at 10% bg efficiency)



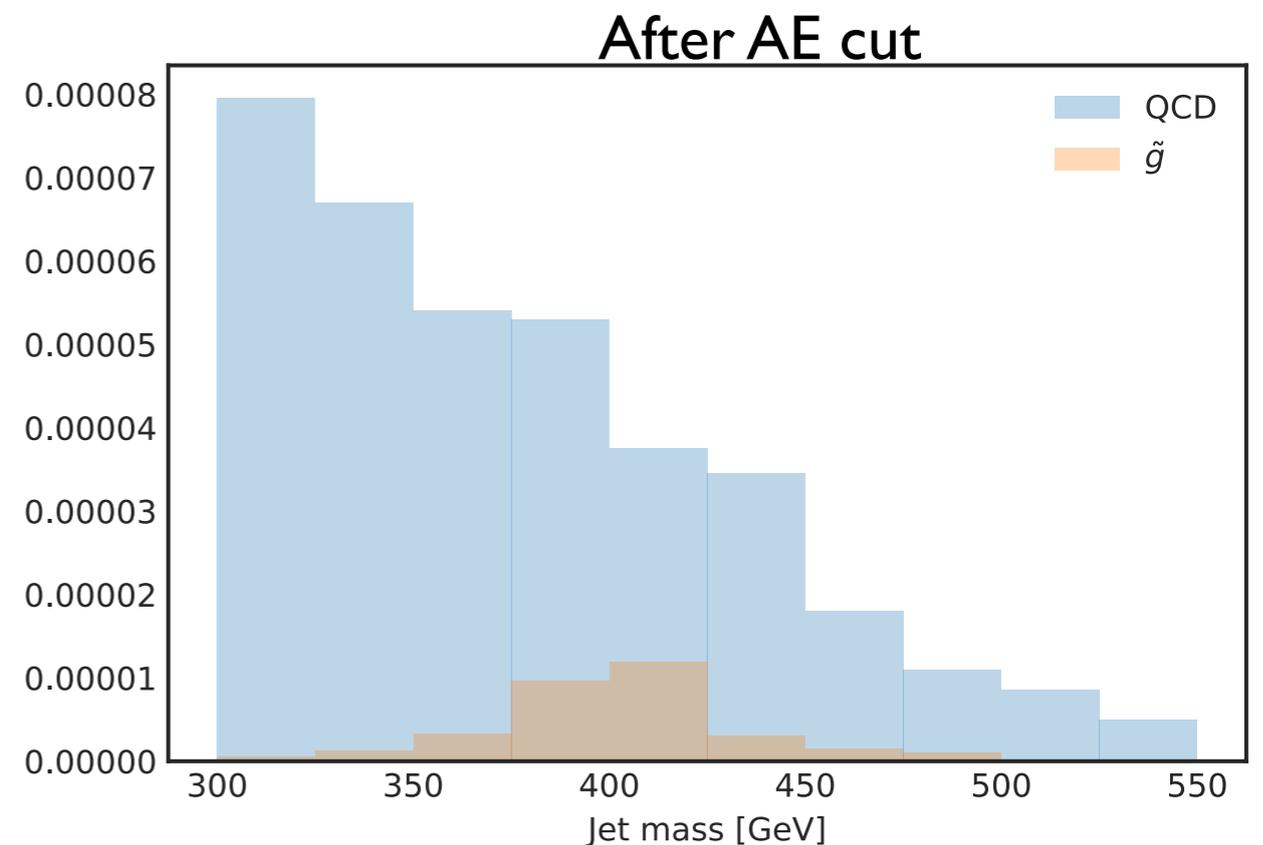
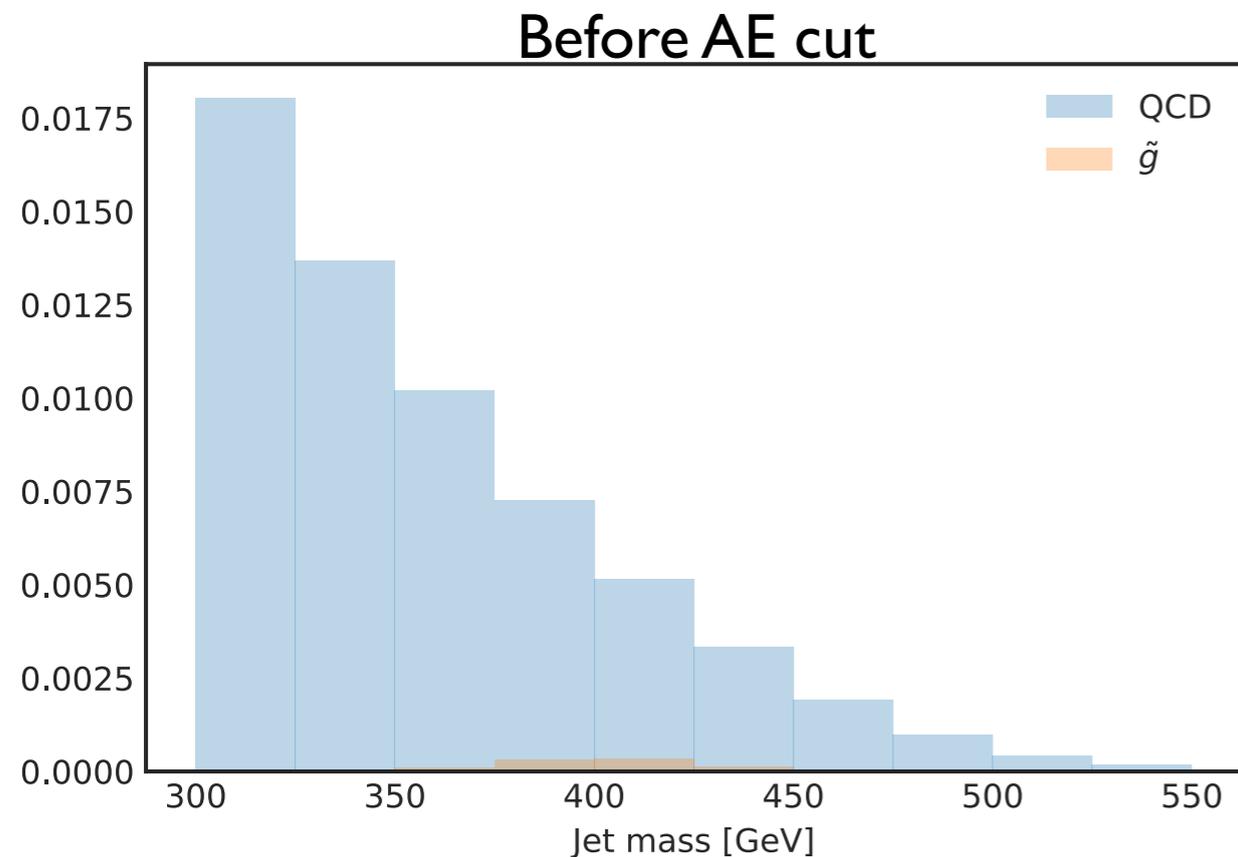
Performance of AE robust even up to 10% contamination!

Autoencoder with explicit decorrelation

Heimel et al 1808.08979



Bump hunt with autoencoder



Train directly on data that contains 400 GeV gluinos.

Use the AE to clean away QCD jets.

Enhance the significance of the bump hunt! (improve S/B by factor of ~ 6)

Could really discover new physics this way!