Search for non-resonant production of semivisible jets with the CMS experiment

Florian Eble

Supervised by Annapaola de Cosa, Jeremi Niedziela and Roberto Seidita

Zurich PhD seminar 2022

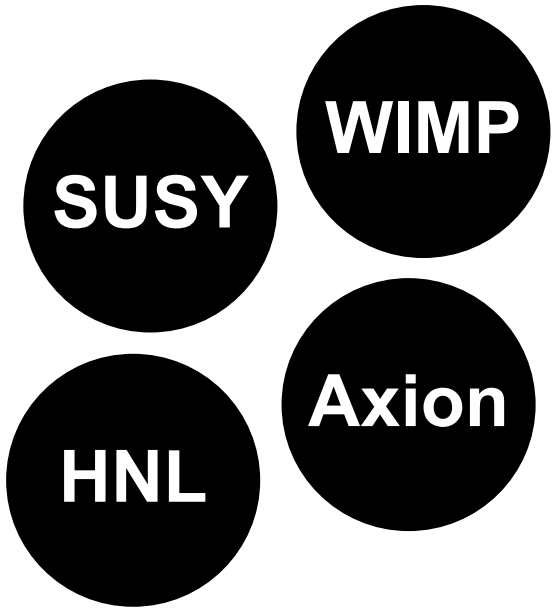
26/01/2023

Dark Matter at a glance

What Dark Matter (DM) is:

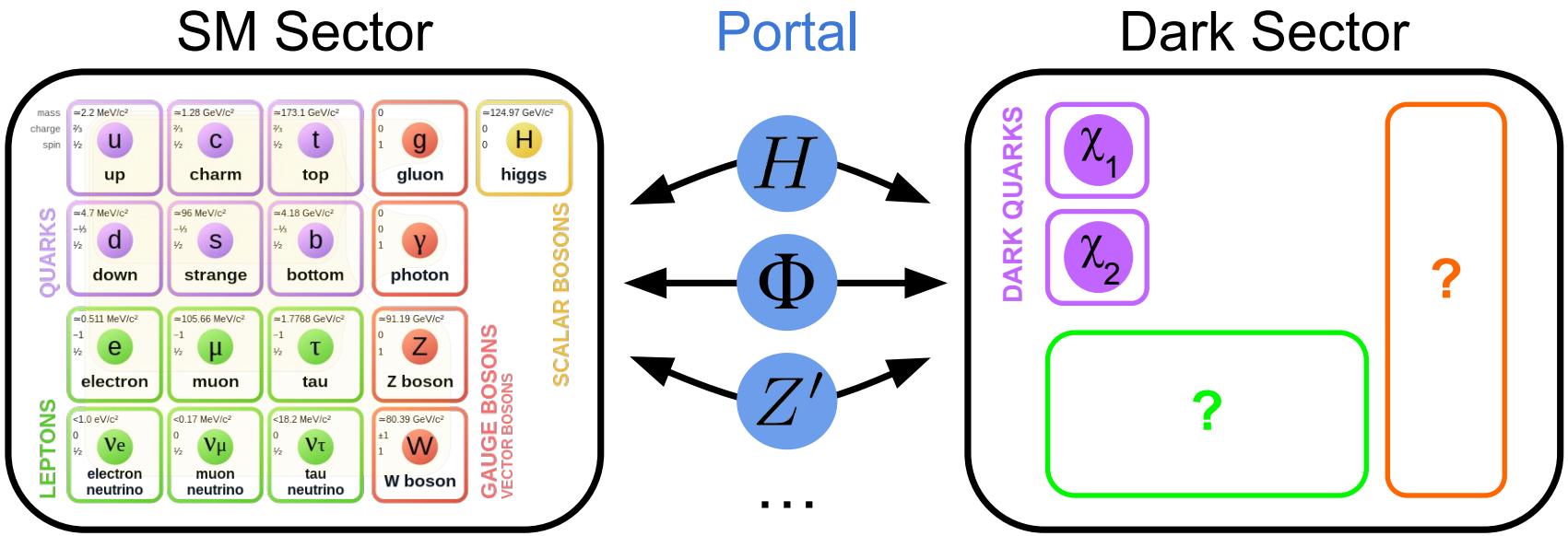
- In practice: Anything that is not Standard Model (SM)
- Experimentally: Non-visible matter that interacts gravitationally (astronomical observation)
- Theoretically: Pick your poison! Supersymmetry (SUSY), Weakly Interacting Massive Particles (WIMPs)...

What if we are not looking at the right place?



DM as a strongly coupled dark sector

- Hidden Valley (arXiv:hep-ph/0604261) with new particles and forces form the dark sector
- There could exist a new confining $SU(N)$ force (a.k.a. dark QCD) and dark quarks
- Mediator particle makes a portal between the SM and dark sectors



Motivations

- Can be probed with collider experiment
- Signatures unexplored by WIMP searches
- Stable dark hadrons could explain the DM relic abundance! (arXiv:1907.04346)

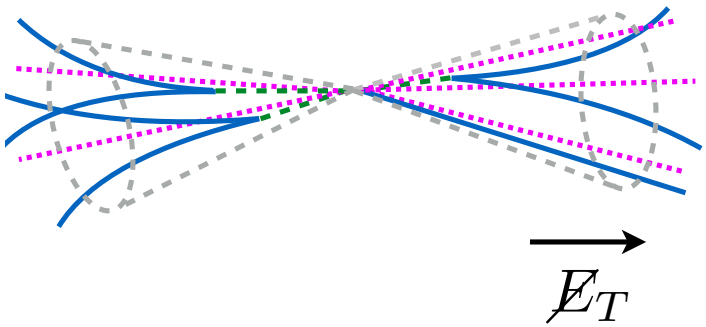
Production of semivisible jets

- Dark quarks hadronize in the dark sector
- A fraction of dark hadrons promptly decays to SM quarks which hadronize in the SM sector
- Remaining dark hadrons are stable and invisible $\implies$ DM candidates

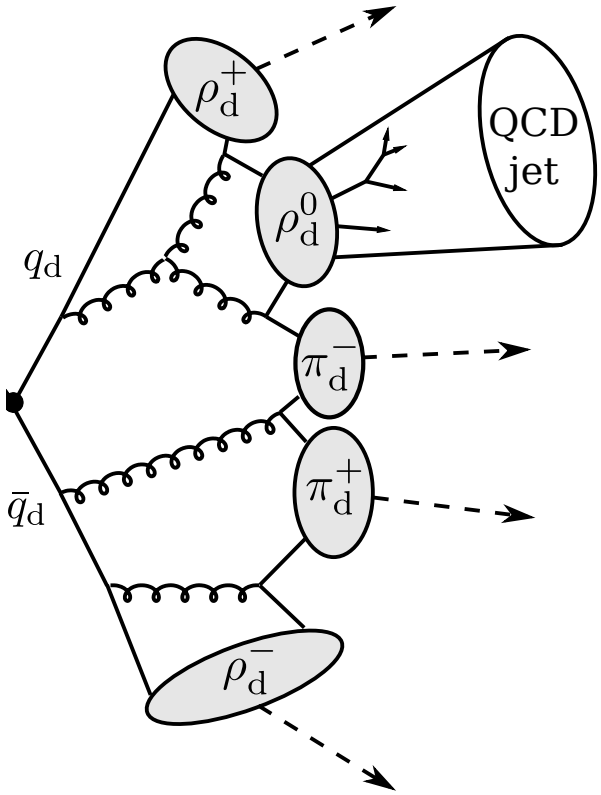
➔ Production of semivisible jets (SVJ) (arXiv:1503.00009, arXiv:1707.05326)

➔ $\not{E}_{\mathbf{T}}$ **aligned with jet!**

$$\not{E}_{\mathrm{T}} = \left|\left|\sum \vec{p_{\mathrm{T}}}\right|\right|$$



SM hadrons
Stable dark hadrons

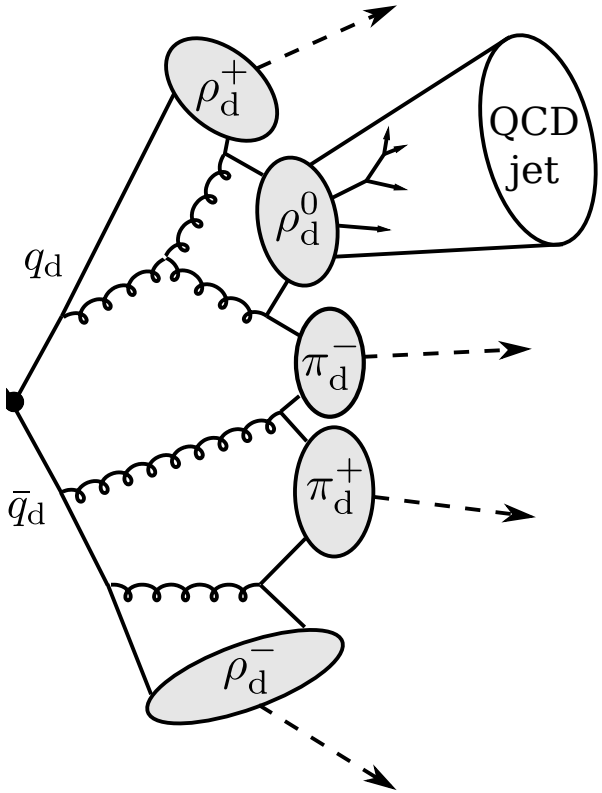


The details of the shower in the dark sector depends on many unknown parameters, e.g.:

- Number of colors and flavors in the dark sector
- Masses of the dark hadrons
- Dark QCD hadronization scale

?

➔ Simulation of SVJ very assumption- and model-dependent



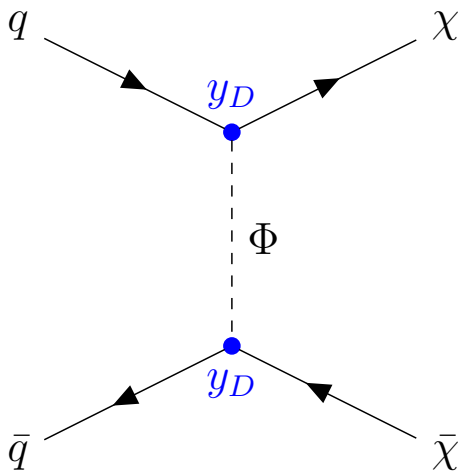
3 production mechanisms:

- **Direct production**:
  Production of dark quarks without resonance
- **Associated production**:
  Production of the mediator associated with a dark quark
- **Pair production**:
  Production of a pair of mediators

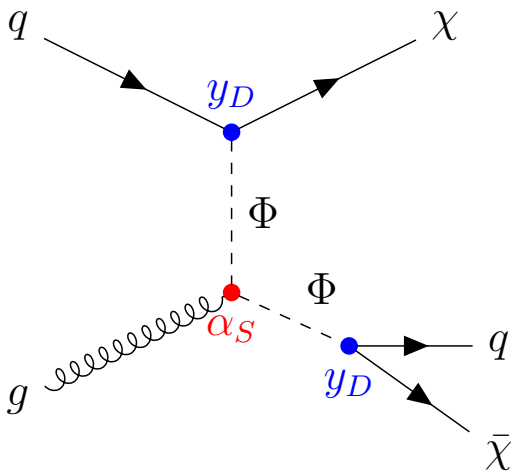
Main model parameters:

- m_Φ: Mass of the mediator Φ
- r_{inv}: Jet invisible fraction

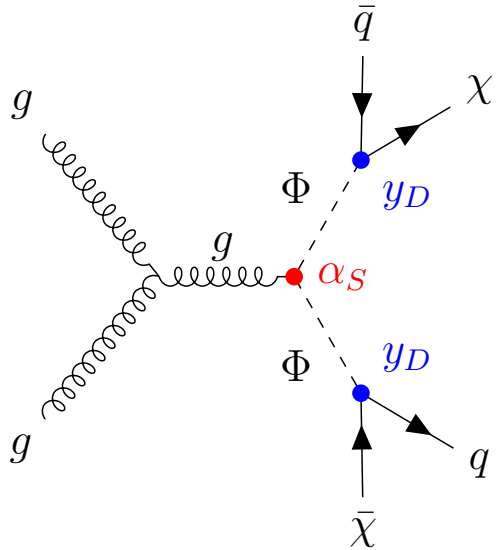
$$r_{\text{inv}} = \left\langle \frac{\text{Number of stable dark hadrons}}{\text{Number of dark hadrons}} \right\rangle$$



(a) Direct production



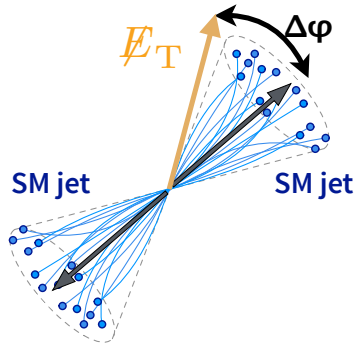
(b) Associated production



(c) Pair production

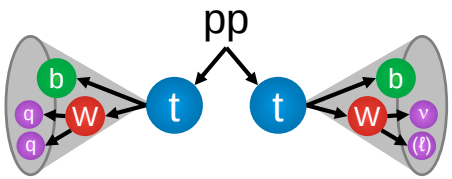
QCD multijet

- Artificial missing transverse energy $\not{E}_{\mathrm{T}}$ aligned with jet from jet energy mismeasurement
- Large cross-section



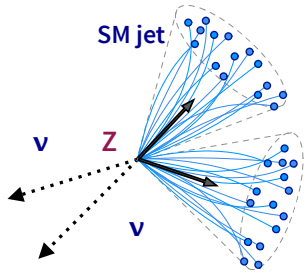
$\mathbf{t\bar{t}}$

- Jet from boosted t
- Semi-leptonic channel $W(\rightarrow l\nu)$ with lost lepton, genuine $\not{E}_{\mathrm{T}}$ from neutrino
- Jet aligned with $\not{E}_{\mathrm{T}}$



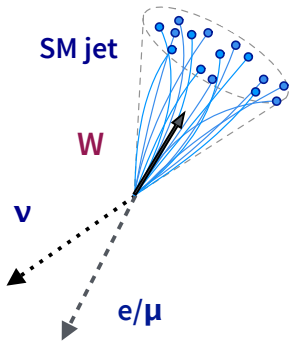
Z + jets

- Genuine $\not{E}_{\mathrm{T}}$ from $Z \rightarrow \nu\nu$



W + jets

- $W \rightarrow l\nu$ with lost/not reconstructed lepton or hadronic decay of τ
- Genuine $\not{E}_{\mathrm{T}}$ from neutrino

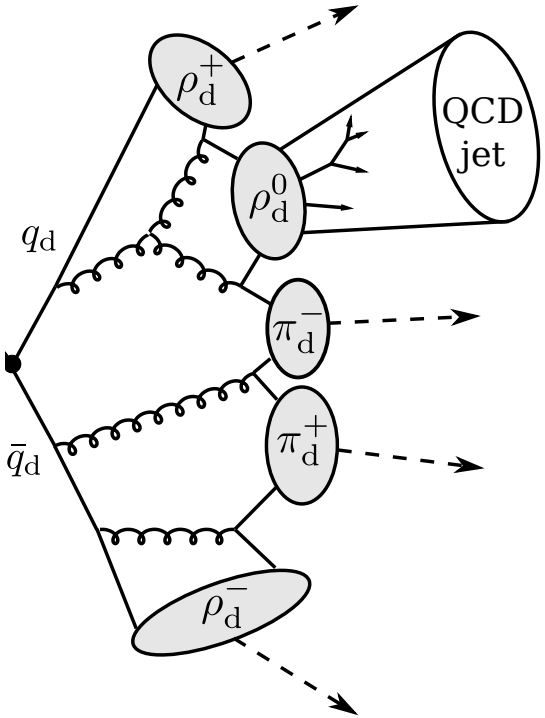


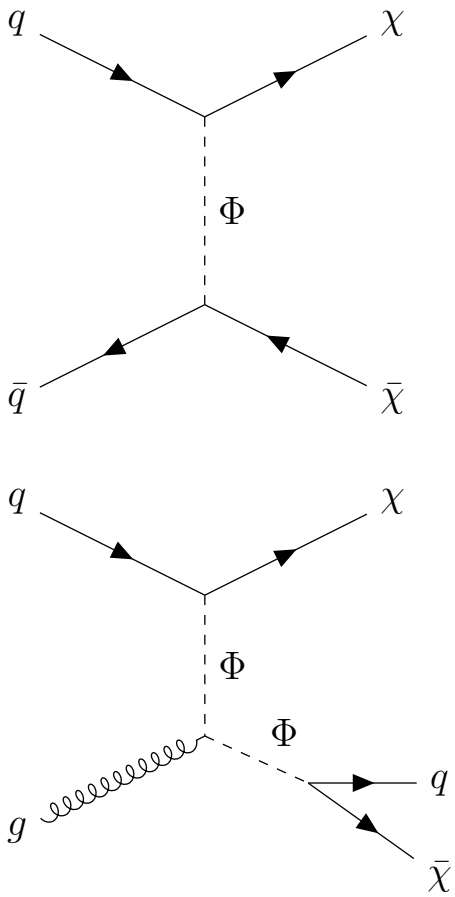
Overall strategy

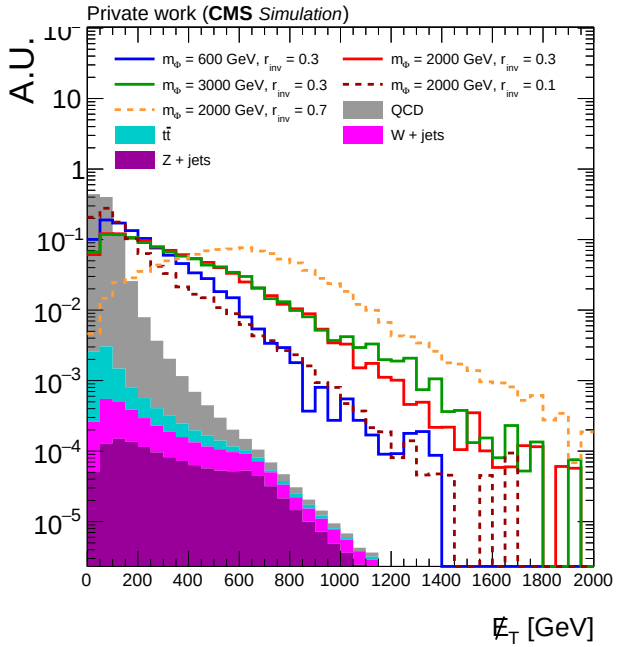
Tag semivisible jets

Classify signal vs background

Search for an excess at high $\not{E}_{\mathrm{T}}$







Challenges

Background

- Overwhelming QCD multijet background in the region of interest ($\not{E}_{\mathrm{T}}$ aligned with jet)
- Potentially several kinds of background jets:
  - light flavor QCD jets
  - boosted 3-prong top jets
  - boosted 2-prong W jets
  - b jets
  - whose relative proportions depend on event-selection

Detector effects

- Energy mismeasurement in calorimeter could mimic the SVJ signature

  e.g. SM dijet events with one jet falling in a calorimeter "cold" cell

Signal modeling

- Non perturbative QCD theory parameters obtained from measurements

  ➔ Not possible for dark QCD!

- Model-dependence:

  Exploiting details of the SVJ shower simulation $\Longrightarrow$ high model-dependence $\Longrightarrow$ small sensivity if the actual dark QCD is different from the one simulated

Pre-selection

Pre-selection used to loosely select a region of phase-space enriched in SVJ events.

Cut-flow table (relative cut efficiencies in %):

m_Φ [GeV]	1000	2000			4000	QCD[1]	$t\bar{t}$	W+jets[2]	Z+jets[2]
r_{inv}	0.3	0.1	0.3	0.7	0.3				
Trigger	32.4	15.0	20.7	21.9	21.9	1.36	6.76	29.6	39.0
$\not{E}_{\text{T}}$ filters	100	99.9	100	100	99.9	99.6	99.9	99.9	99.9
Lepton Veto	93.0	92.1	92.7	95.0	93.4	91.9	35.5	36.1	94.3
$S_{\text{T}} > 1300$ GeV	37.7	58.3	43.6	31.9	26.5	7.58	9.34	3.95	1.31
nFatJet ≥ 2	99.9	100	99.7	95.5	99.6	99.9	99.9	97.7	91.0
Total abs. eff. [%]	11.3	8.07	8.36	6.34	5.39	0.11	0.24	0.41	0.44

- $\not{E}_{\text{T}}$ filters aim at rejecting events with artificial $\not{E}_{\text{T}}$ caused by detector mismeasurement, like calorimeter "cold" cell
- Trigger + S_{T} (trigger plateau) cuts have low signal efficiency!
- Lepton veto efficient at rejecting $t\bar{t}$ and W + jets backgrounds

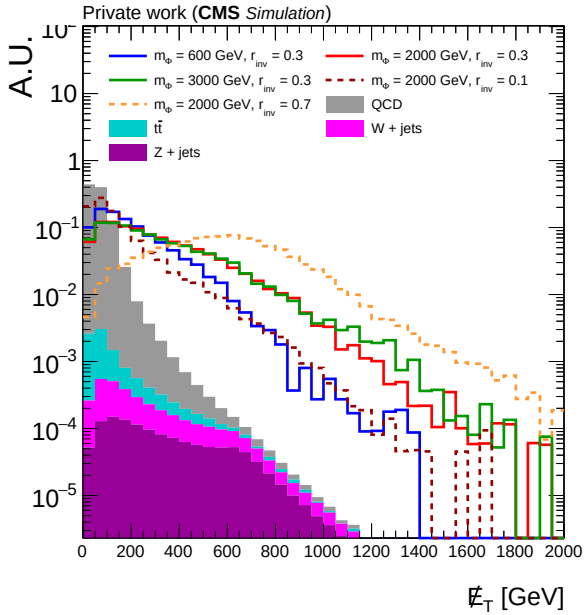
[1] Gen-level $\hat{p_{\text{T}}} > 170$ GeV

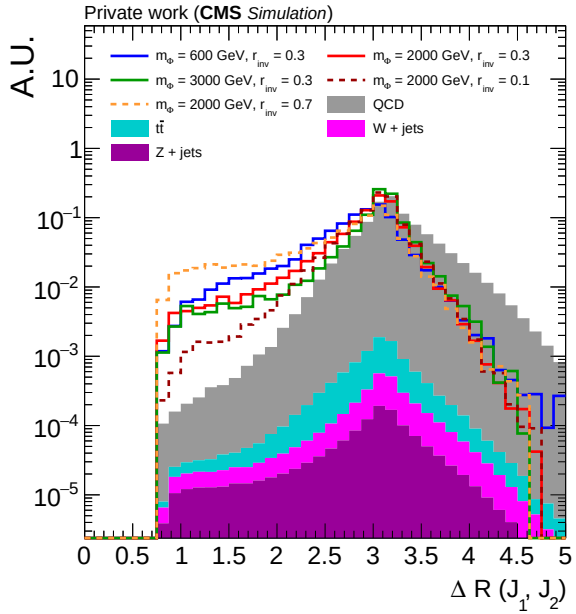
[2] Gen-level $H_{\text{T}} > 400$ GeV

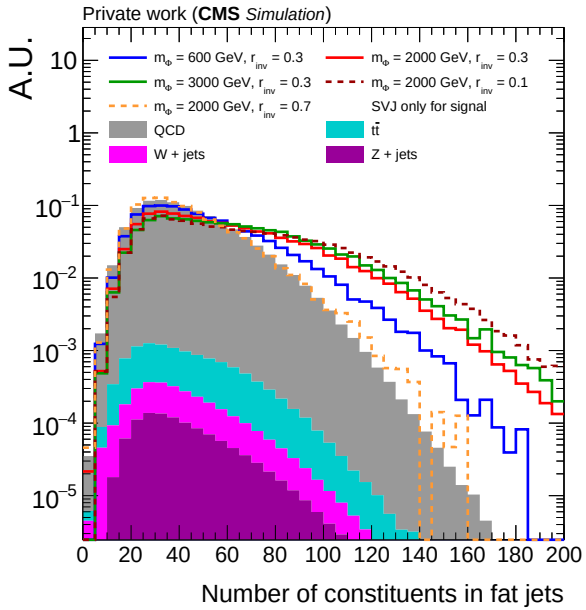
Overview of kinematics and jet substructure

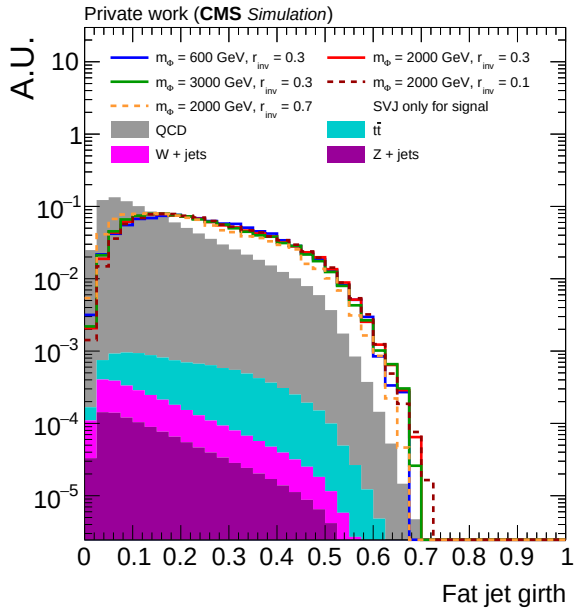
- $\not{E}_{\mathrm{T}}$ is the most discriminative event-level variable
- Moderate discrimination in other event-level variables, *e.g.* angular variables, usually enhanced at high r_{inv}
- Moderate discrimination in various jet substructure (JSS) variables, usually enhanced at low r_{inv}

➔ JSS and event-kinematics have complementary sensitivity









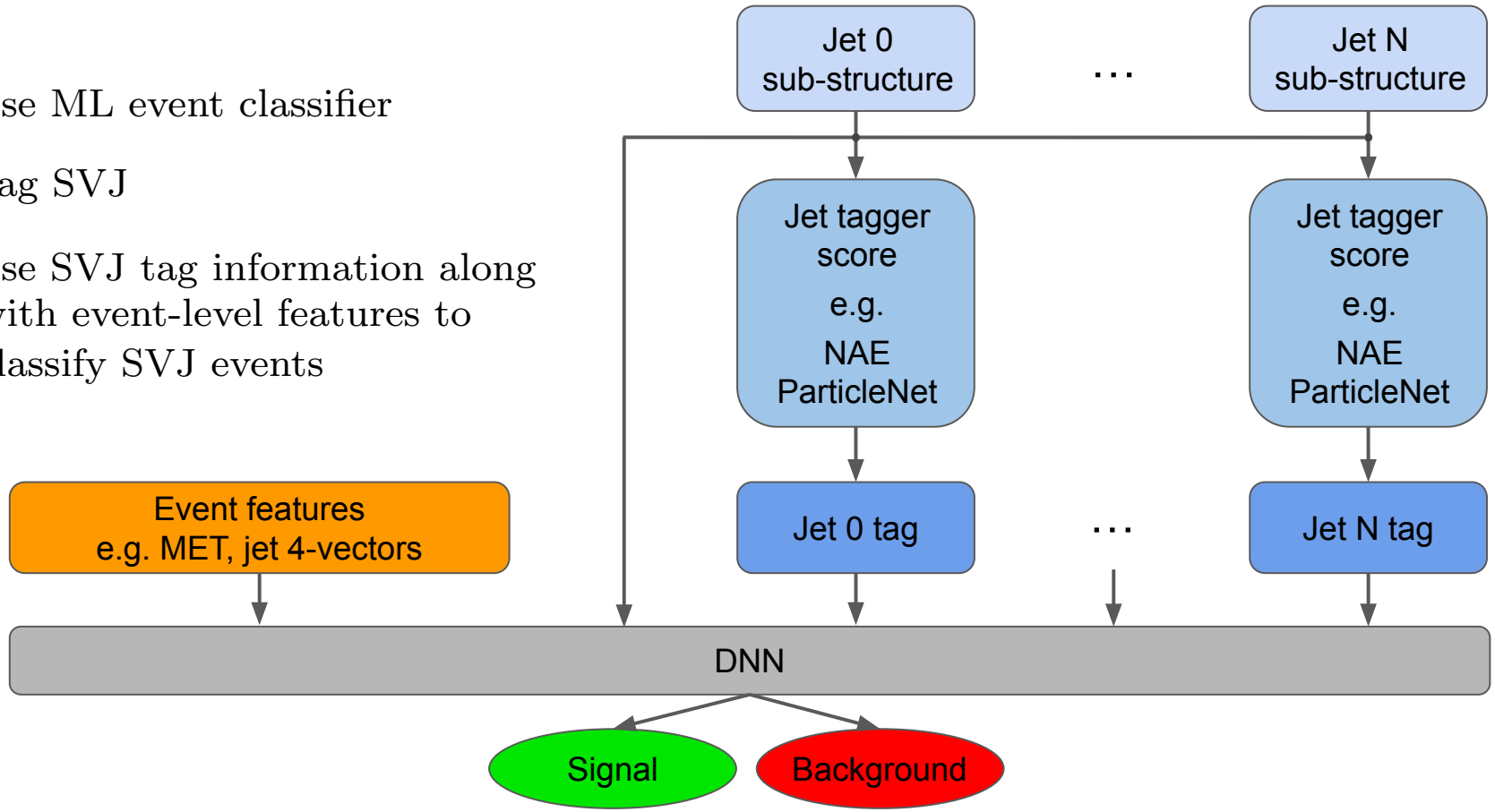
Analysis strategy

- Several event-level variables contain moderate discrimination power
- No single obvious discriminative variable for all model parameters and production mechanisms

➔ Use ML event classifier

➔ Tag SVJ

➔ Use SVJ tag information along with event-level features to classify SVJ events



Analysis strategy - Tag semivisible jets/events

3 approaches:

Jet tagger \ Event classifier	Unsupervised	Supervised
Unsupervised	**Fully unsupervised**	**Mixed**
Supervised		**Fully supervised**

Supervised vs unsupervised jet tagger:

- Large model dependence for dark shower simulation
- Unsupervised SVJ tagger can achieve good discrimination power and surpasses supervised tagger in case of test on an unseen model (arXiv:2112.02864)
- Unsupervised tagger can be trained directly on data in a control region!

Supervised vs unsupervised event classifier:

- Event kinematics is better simulated and less model-dependent
- Supervised classifier: boost sensitivity compared to unsupervised approach

Autoencoders

Autoencoders (AE) are neural networks (NN) classically used for dimensionality reduction or anomaly detection.

AEs are composed of:

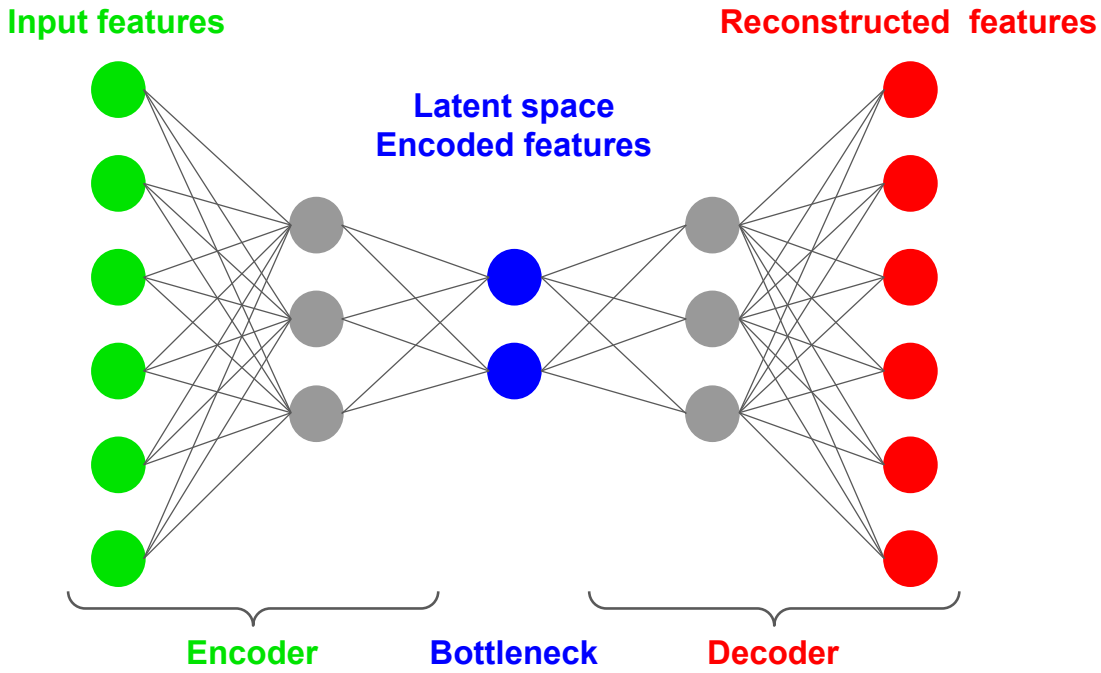
- an encoder NN f
- a "symmetric" decoder NN g

The AE network is trained to learn to reconstruct the input examples it is given.

The loss of an AN for an example x is:

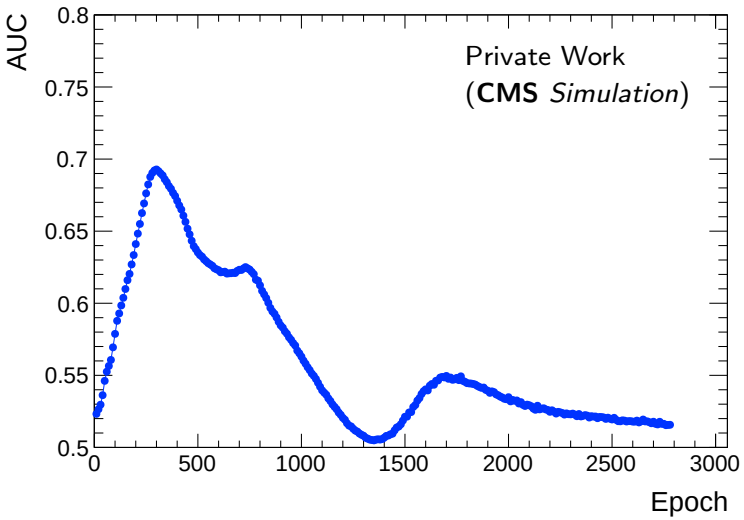
$$L(x) = ||g(f(x)) - x||$$

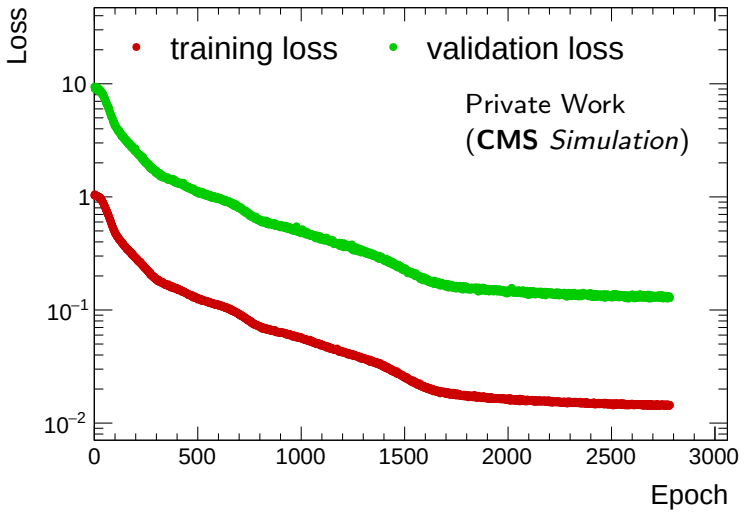
where $||\cdot||$ is a distance



The aim of an AE for anomaly detection is to be able to reconstruct only the examples it is trained on but not others

Shortcoming of standard autoencoder



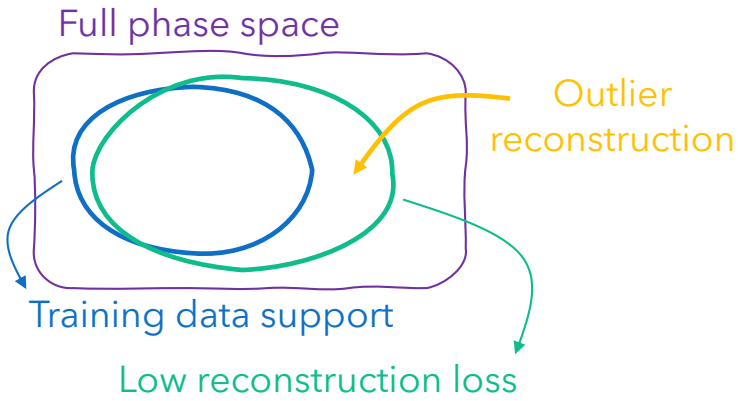


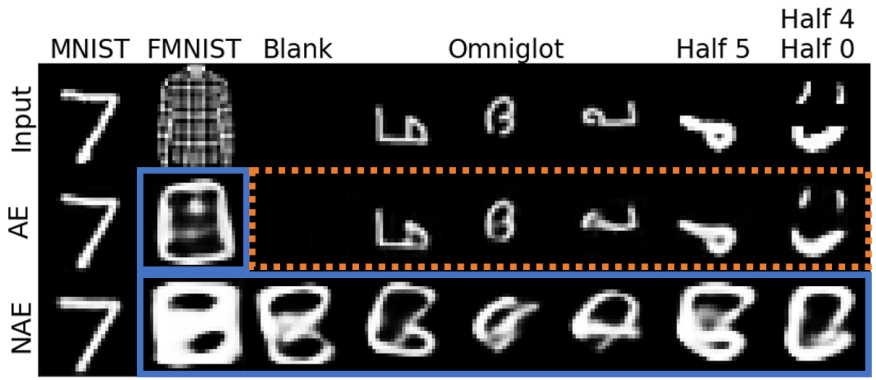
- Training plain autoencoder (AE) on $t\bar{t}$ jets
- 2 phases...
  - (A) Increasing AUC and decreasing reconstruction loss
  - (B) Decreasing AUC while the reconstruction loss continues to decrease
- ... corresponding to the AE
  - (A) getting better at reconstructing background jets but not signal jets as much
  - (B) getting better at reconstructing signal jets, as good as background jets
    - ➔ Outlier reconstruction / AE is $\approx$ identity!
- Rather large AUC (≈ 0.7), however for a very particular point during training

➔ Ultimately, final AUC is only slightly above 0.5

[1] Training and validation loss on different scale because training loss is divided by number of input features while validation loss is not

The problem of outlier reconstruction





Outlier reconstruction example: AE and NAE trained on MNIST, other inputs are outliers.

- Outlier reconstruction happens when the network assigns low reconstruction error to out-of-distribution (OOD) examples
- OOD reconstruction not suppressed during training in plain AE
- Sometimes phrased as "OOD examples need to be more 'complex' to not be reconstructed"

➔ **Normalized autoencoder[1] (NAE) features a mechanism to suppress OOD reconstruction!**

➔ It ensures that the low error phase-space of the NN matches that of the training data.

[1] NAE first introduced in arXiv:2105.05735 and used in HEP in arXiv:2206.14225

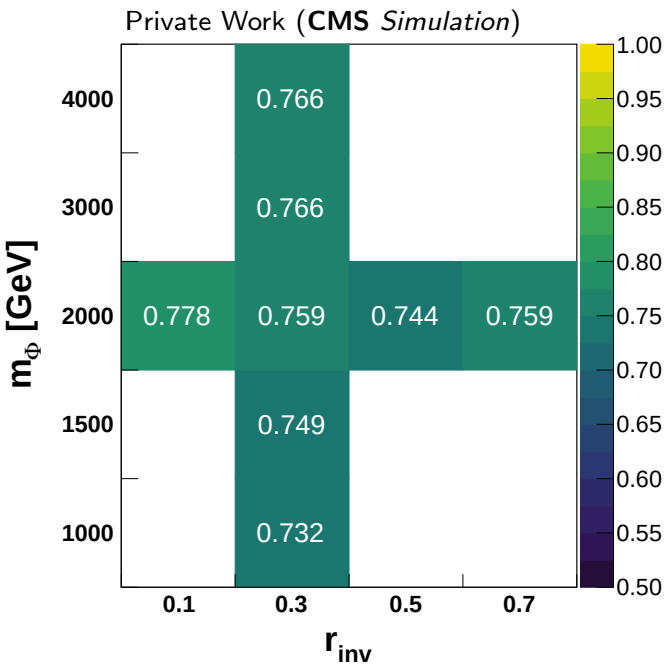
NAE against $t\bar{t}$ and QCD separately

Best model choice:

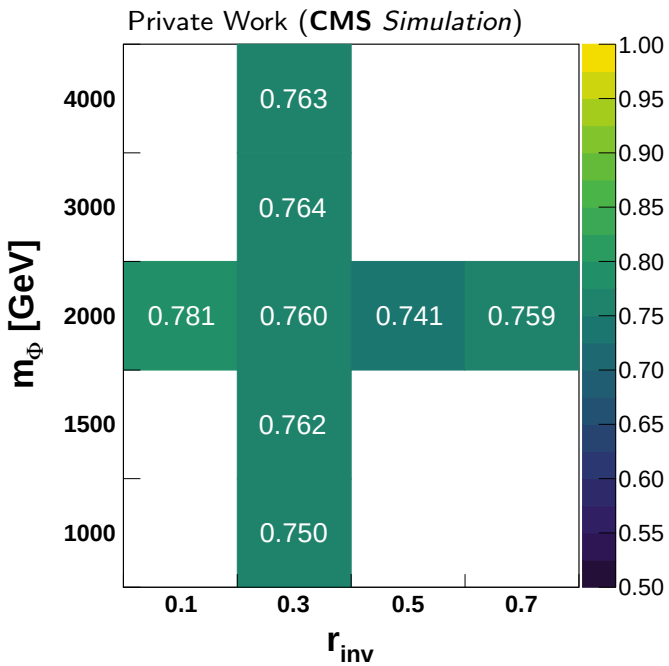
- Monitoring AUC calculated on an ensemble of signal hypotheses
- Choosing model with highest AUC during training

Successful unsupervised SVJ tagger against QCD and $t\bar{t}$ jets!

$t\bar{t}$ vs SVJ



QCD vs SVJ



SVJ events selection

SVJ events selection:

- Exploit weakly discriminative variables with DNN (*e.g.* angles between jets and with $\not{E}_{\mathrm{T}}$)
- Signal region: high DNN and high $\not{E}_{\mathrm{T}}$
- DNN output must be uncorrelated with $\not{E}_{\mathrm{T}}$ to perform background estimation with ABCD method (see slide 19)

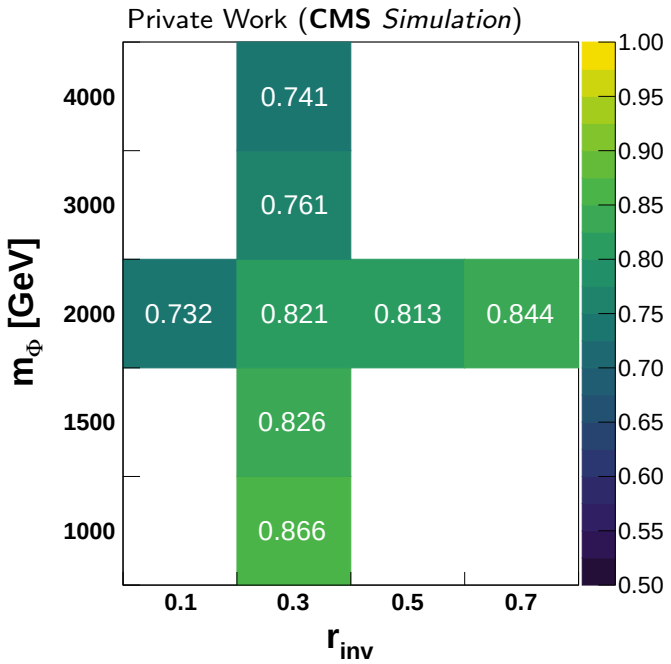
Example:

- Trained DNN using η, ϕ of first 4 leading large[2] jets and $\not{E}_{\phi}$
- Used 0-zero padding if fewer than 4 jets
- Trained on a mixture of all main backgrounds

➔ Achieving large AUC!

Uncorrelation between DNN score and $\not{E}_{\mathrm{T}}$:

- Either by providing input features uncorrelated with $\not{E}_{\mathrm{T}}$, usually weakly discriminative
- Or by using also features correlated with $\not{E}_{\mathrm{T}}$ and decorrelating DNN score with $\not{E}_{\mathrm{T}}$ using DisCo[1]



[1] The distance correlation (DisCo) is a measure of non-linear correlation between two variables. See backup slide 18.

[2] Reconstructed with the anti-k_{T} algorithm with radius $R = 0.8$

Background estimation and statistical analysis

Background estimation:

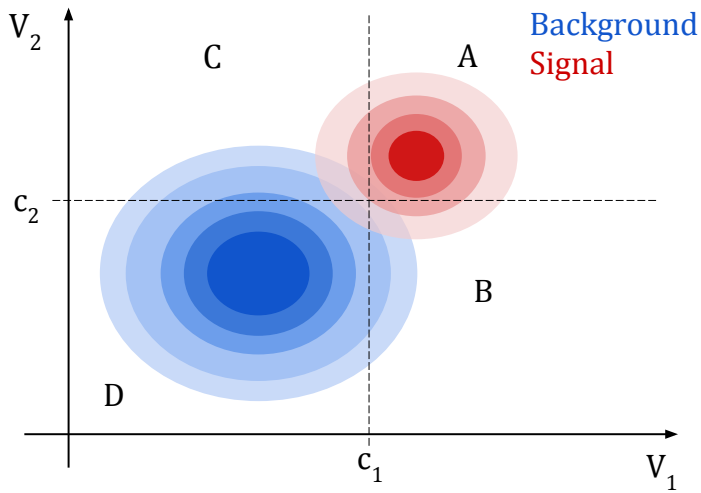
- Data-driven technique as QCD multijet simulation is not reliable
- ABCD method: event-classifier score versus an uncorrelated variable (e.g. $\not{E}_{\mathrm{T}}$)

Statistical analysis:

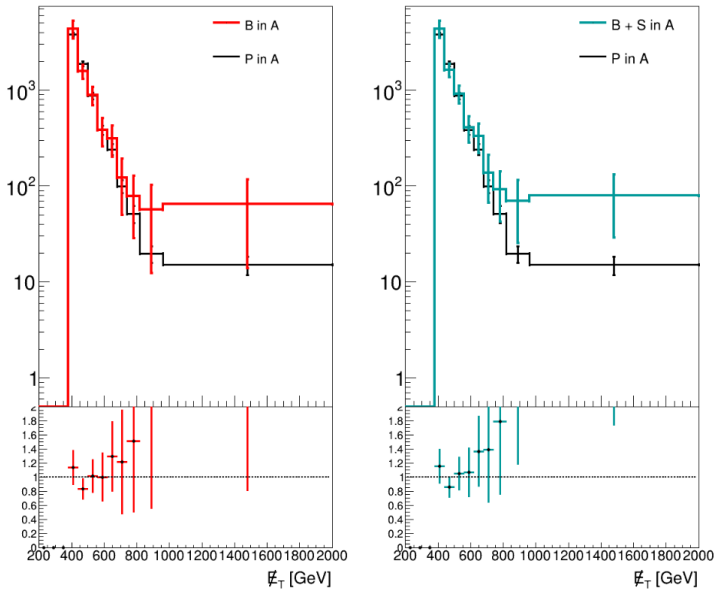
- Search for an excess in $\not{E}_{\mathrm{T}}$ in the signal region

Background estimation in signal region:

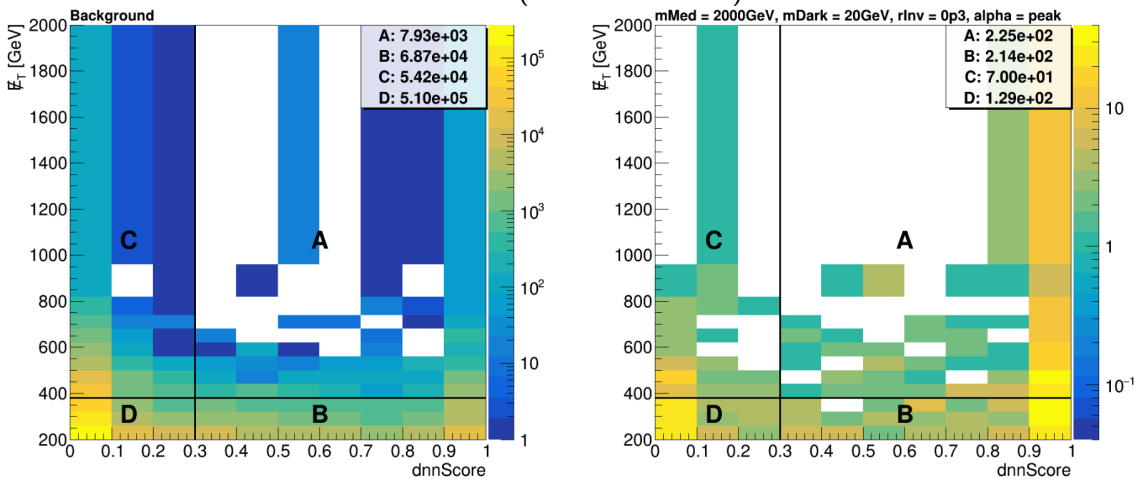
$$N_A^{\mathrm{bkg}} = \frac{N_B N_C}{N_D}$$



Private Work (**CMS** *Simulation*)



Private Work (**CMS** *Simulation*)



Conclusions and next steps

Current status:

- Analysis strategy well in place
- Proof-of-principle for unsupervised jet tagger (NAE) selecting SVJ against a mixture of SM background jets
- Supervised event classifier (DNN) exploiting weakly discriminative variables or more discriminative variables with DisCo
- Proof-of-principle for data-driven background estimation using ABCD method

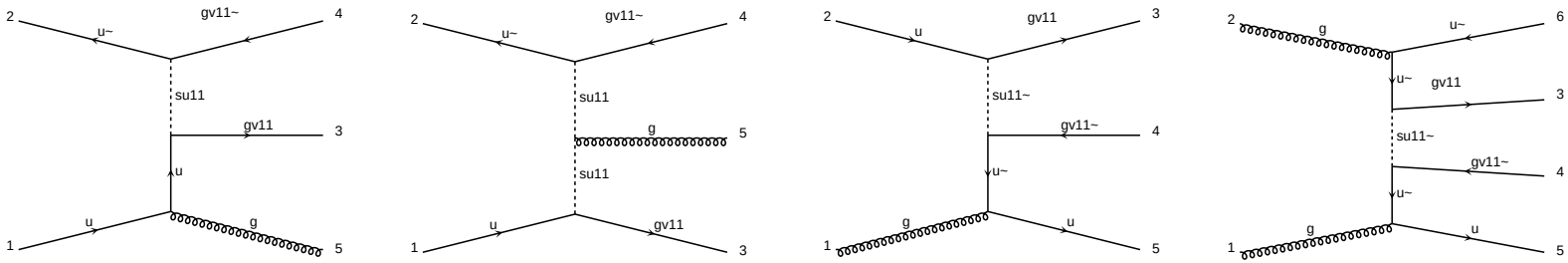
Next steps:

- Finalize jet tagger:
  - Optimization of input features
  - Optimization of hyper-parameters with Optuna
  - Checking correlation with high level jet features, *e.g.* p_{T}
- Finalize event classifier:
  - Add NAE loss to input feature
  - Finalize input features in connection with uncorrelation with $\not\!\!E_{\mathrm{T}}$
- Verify/optimize artificial $\not\!\!E_{\mathrm{T}}$ filters for this channel
- Perform statistical analysis including systematics

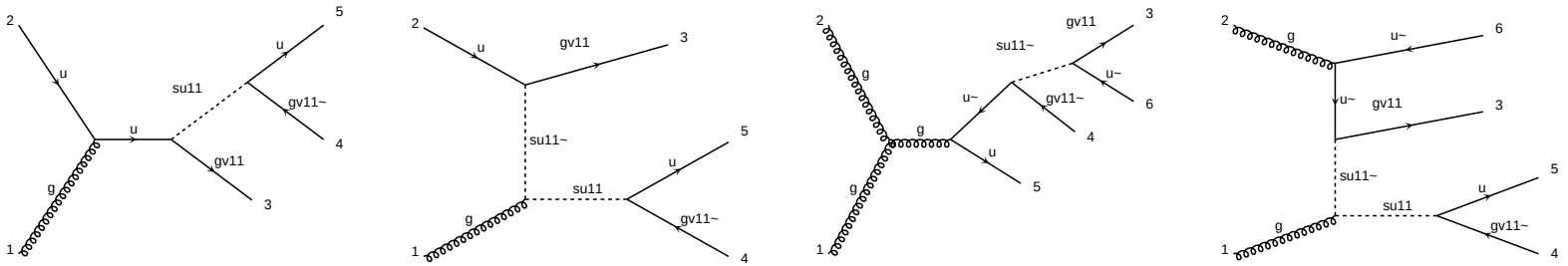
Backup

Many possible diagrams in the t-channel

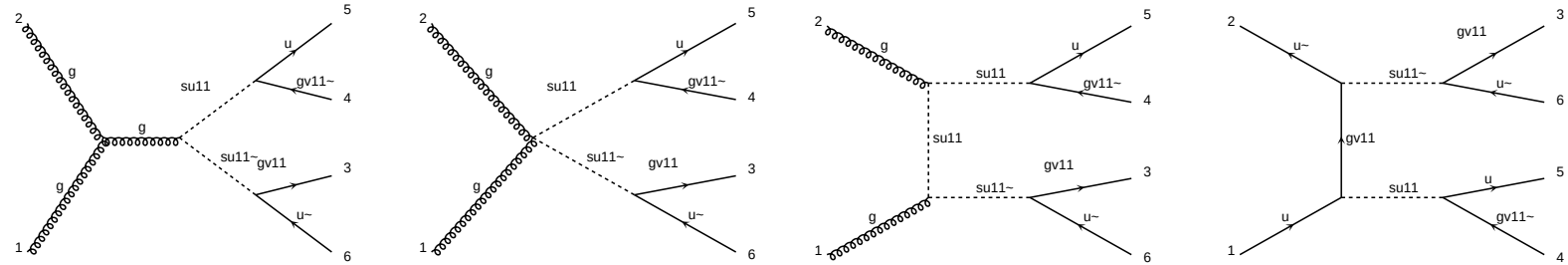
- Direct production



- Associated production



- Pair production



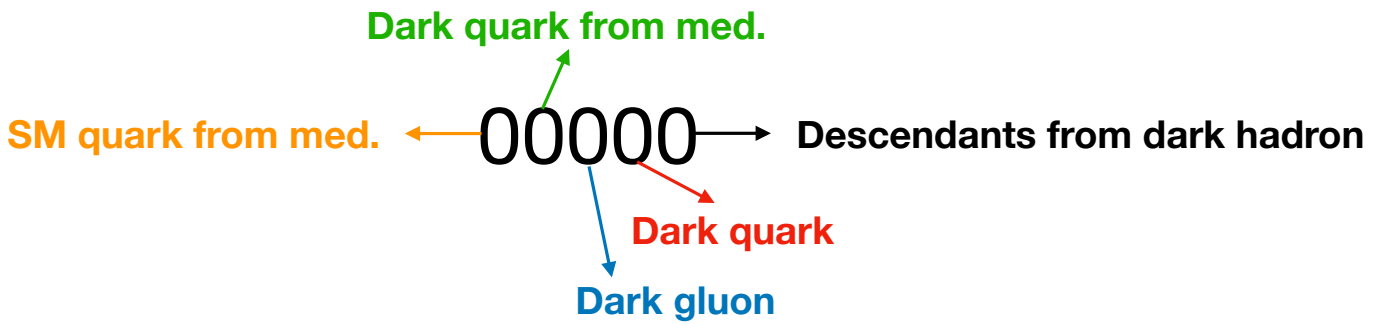
su11 is the mediator Φ, gv11 is a dark quark

t-channel jet classification

Different jets in the t-channel final state:

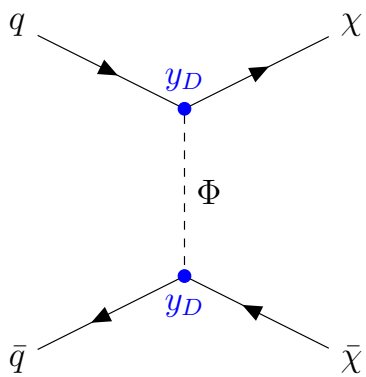
- SVJ not from the mediator
- SVJ from the mediator
- SVJ initiated by a dark quark, dark gluon
- SM jets not from the mediator
- SM jets from the mediator
- Any combination of the previous cases...

Unique code to define each possibility:

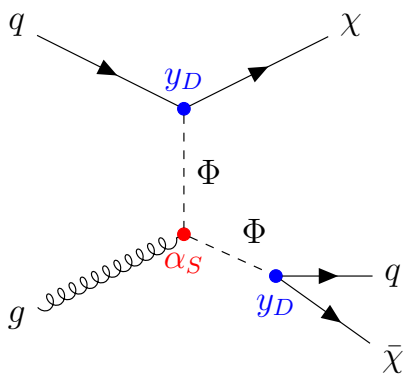


SVJ categories:
1 - 15

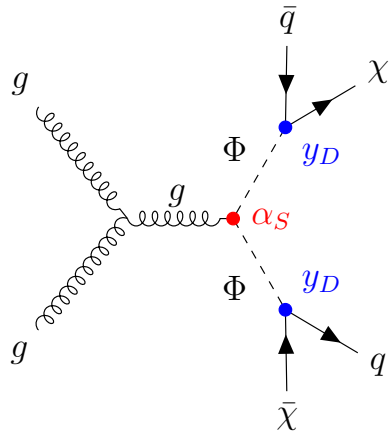
Jets from dark quark, dark gluon, from the mediator or not



(a) Direct production



(b) Associated production



(c) Pair production

Energy-based models (EMBs)

- EBMs are models where the probability is defined through the Boltzmann distribution
- Let θ denote the model parameters
- The model probability p_θ is defined from the energy E_θ

$$p_\theta(x) = \frac{1}{\Omega_\theta} \exp\left(-E_\theta(x)/T\right) \tag{1}$$

where the normalization constant Ω_θ is

$$\Omega_\theta = \int \exp\left(-E_\theta(x)/T\right) dx \tag{2}$$

- The EBM loss for a training example x is the negative log-likelihood:

$$L_\theta(x) = -\log p_\theta(x) = E_\theta(x)/T + \log \Omega_\theta \tag{3}$$

- The gradient of the EBM loss is thus:

$$\nabla_\theta L_\theta(x) = \nabla_\theta E_\theta(x) - \mathbb{E}_{x' \sim p_\theta}\left[\nabla_\theta E_\theta(x')\right] \tag{4}$$

- The expectation value over the training dataset, with probability p_{data} is:

$$\mathbb{E}_{x \sim p_{\text{data}}}\left[\nabla_\theta L_\theta(x)\right] = \mathbb{E}_{x \sim p_{\text{data}}}\left[\nabla_\theta E_\theta(x)\right] - \mathbb{E}_{x' \sim p_\theta}\left[\nabla_\theta E_\theta(x')\right] \tag{5}$$

Working principle of the Normalized Autoencoder (NAE)

Basic idea:

- **Ensure that low reconstruction error phase-space matches that of training data**
- *i.e.* OOD examples are constrained to have high reconstruction error

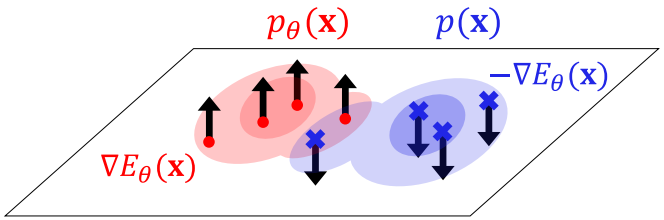


Figure 2. An illustration of the energy gradients in Eq. (7). The red and blue shades represent the model and the data density, respectively. The gradient update following Eq. (7) increases the energy of samples from $p_\theta(\mathbf{x})$ (the red dots) and decreases the energy of training data (the blue crosses).

NAE methematics

- Let θ be the NAE model parameters
- Let p_{data} be the probability distribution of the training data
- Let E_θ be the reconstruction error of the AE
- We define the model probability p_θ through the Boltzmann distribution as:

$$p_\theta(x) = \frac{1}{\Omega_\theta} \exp\left(-E_\theta(x)\right) \tag{6}$$

  Such that the phase space with high reconstruction error has low probability

- We define the loss to learn $p_\theta = p_{\text{data}}$:

$$\mathbb{E}_{x\sim p_{\text{data}}}\left[L_\theta(x)\right] = \mathbb{E}_{x\sim p_{\text{data}}}\left[E_\theta(x)\right] - \mathbb{E}_{x'\sim p_\theta}\left[E_\theta(x')\right] \tag{7}$$

positive energy negative energy

Working principle of the Normalized Autoencoder (NAE)

Loss

$$\mathbb{E}_{x\sim p_{\text{data}}}\left[L_\theta(x)\right] = \mathbb{E}_{x\sim p_{\text{data}}}\left[E_\theta(x)\right] - \mathbb{E}_{x'\sim p_\theta}\left[E_\theta(x')\right] = E_+ - E_-$$

positive energy negative energy

Positive energy

- Simply the reconstruction error over the training dataset
- Take SM jets and compute the reconstruction error!

Negative energy

- Reconstruction error of the "negative samples" x' from the probability distribution p_θ
- Need to sample from the model to get the "negative samples"

➔ Monte Carlo Markov Chain (MCMC) employed

MCMC

- Start from an initial point x'_0
- Run n Langevin MCMC steps:

$$x'_{i+1} = x'_i - \lambda_i \nabla_x E_\theta(x'_i) + \sigma_i \epsilon \qquad \epsilon \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}) \tag{8}$$

drift diffusion

- Repeat with several points $x'^{(j)}_0$, the negative samples are the $x'^{(j)}_n$

Training samples and hyper-parameters

Input features (AK8 jets)

Jet shape	Axis major axis minor,
Jet substructure	$p_{\mathrm{T}}^{\mathrm{D}}$, EFP1
Boosted object	τ_2, τ_3 $C_2^{\beta=0.5}$, $D_2^{\beta=0.5}$
Other	Jet mass[1]

Architecture

Fully connected neural net

Hidden layers: $10, 10, 6, 10, 10$

Hyper-parameters

Hyper-parameter	Value
Batch size	256
Reconstruction loss	MSE
Activation	ReLU
Output encoder/ decoder activation	Linear
Optimizer	Adam
Learning rate	1e-5
Dropout	0.
Max number of epochs	20 000
MCMC	PCD
Regularization	See slide **??**

Number of events

m_Φ [GeV]	1000	1500	2000				3000	4000	QCD	$t\bar{t}$
r_{inv}	0.3	0.3	0.1	0.5	0.3	0.7	0.3	0.3		
Number of events	23k	25k	23k	18k	16k	11k	14k	14k	217k	23k

Number of AK8 jets

Background jets	Leading 2 jets
Signal jets	Only SVJ[2] in leading 2 jets

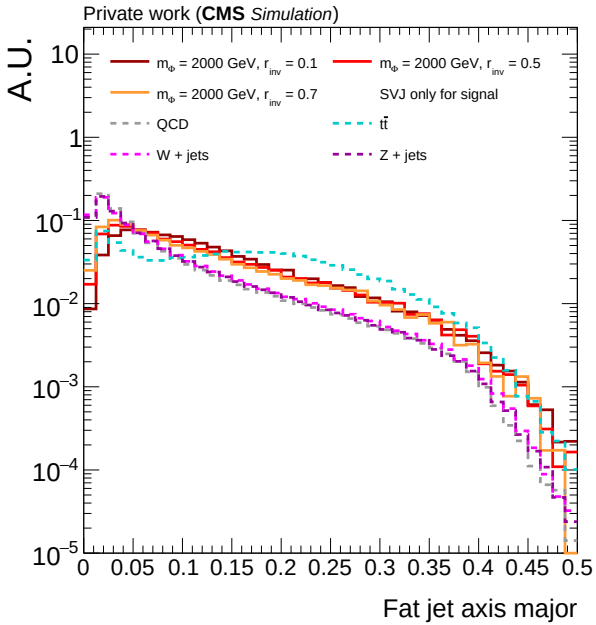
Train/validation/test splitting

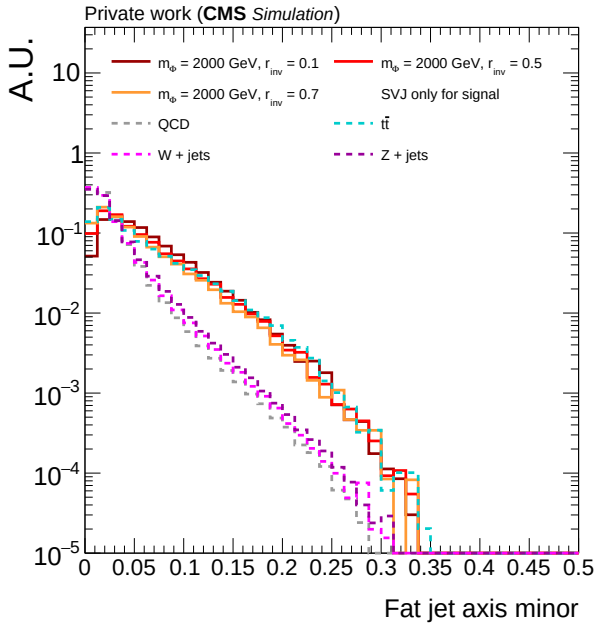
0.7/0.15/0.15

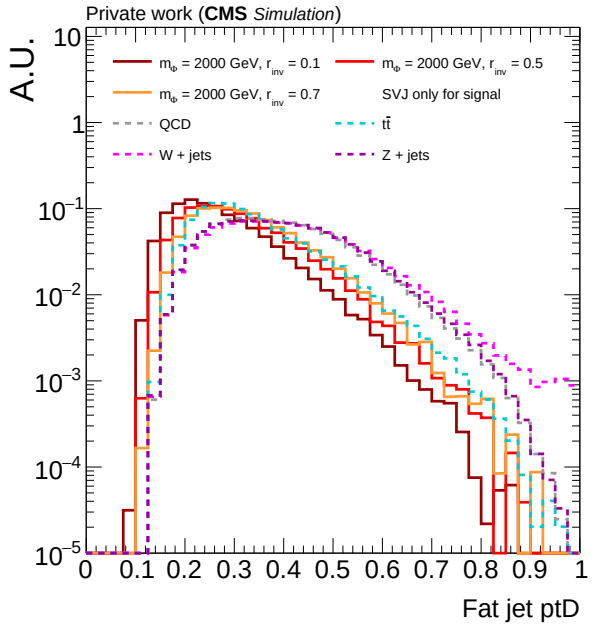
[1] Should use soft drop mass in the future to better exploit dark hadron mass

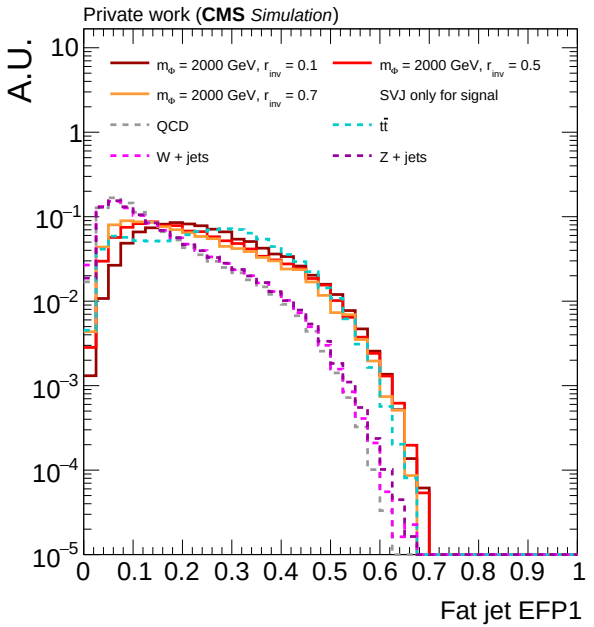
[2] t-channel SVJ, categories explained in backup slide 3

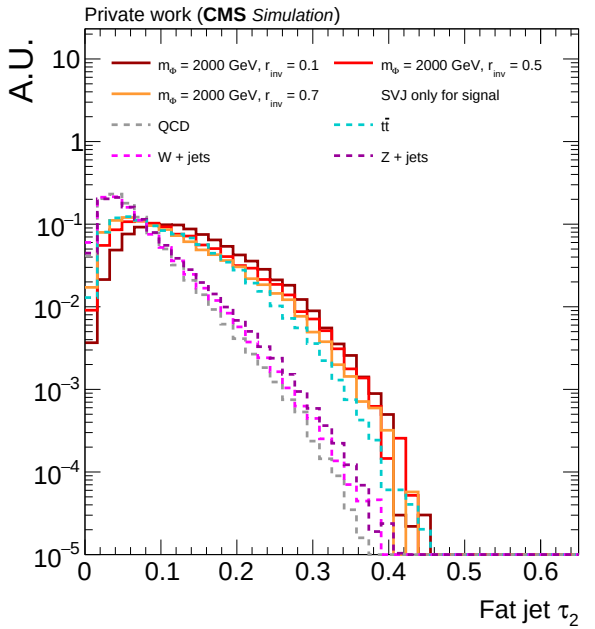
Input features

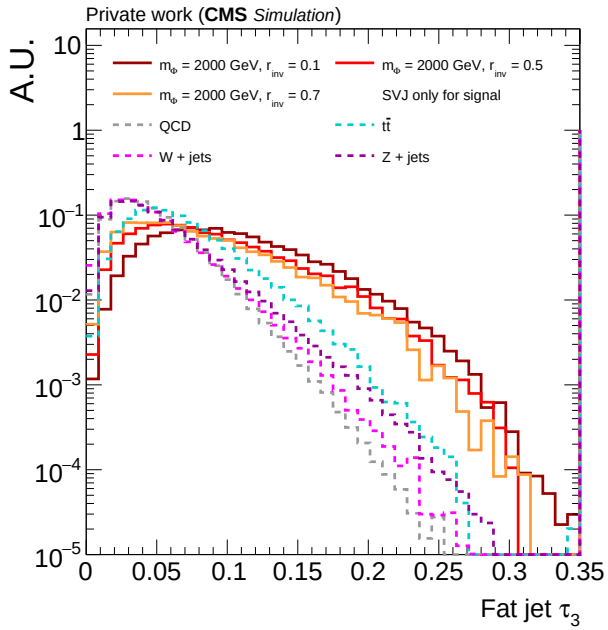




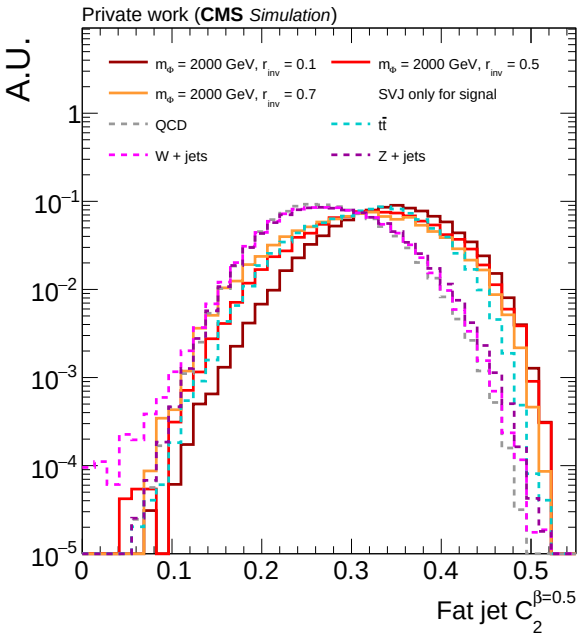


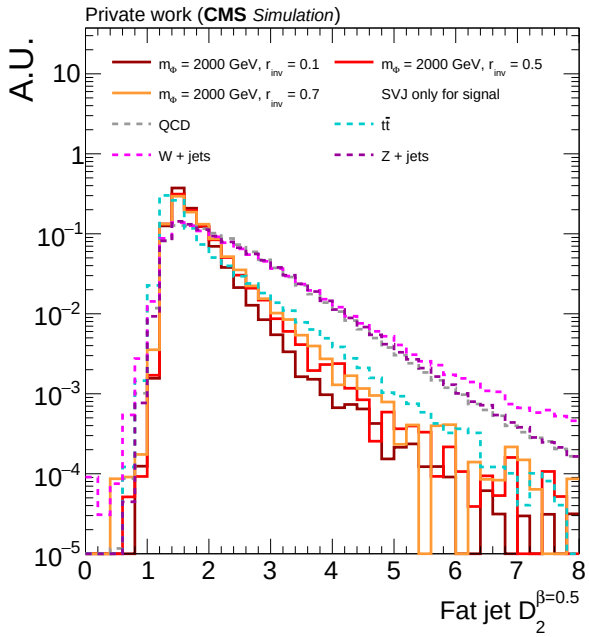


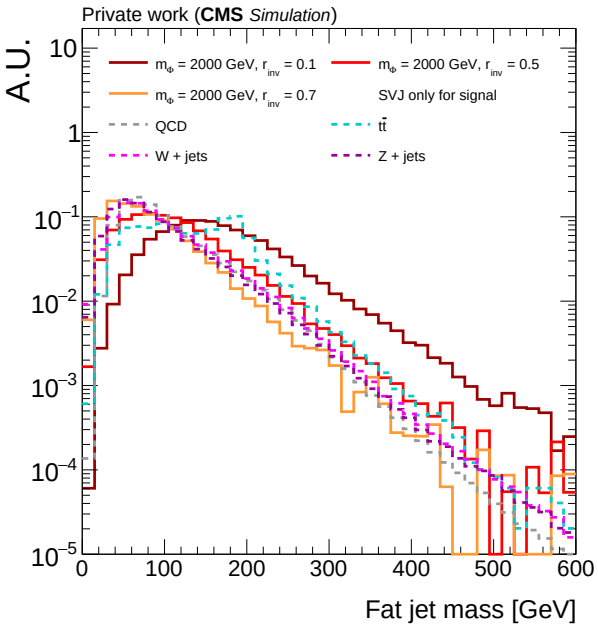


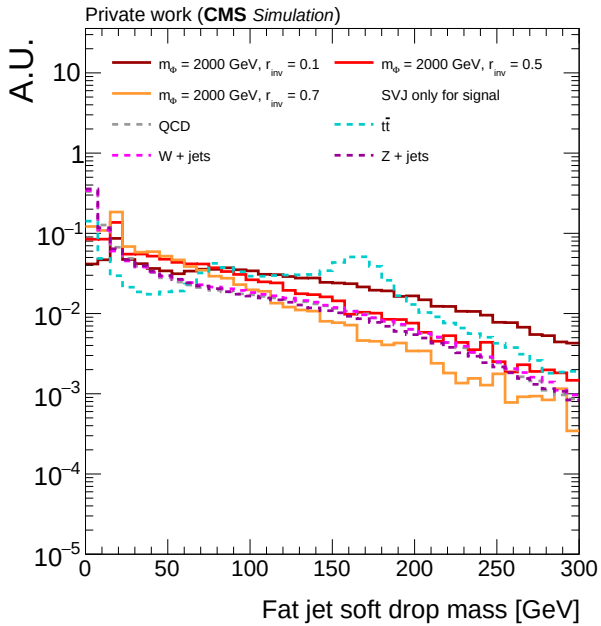


Input features









In general:

- SVJ JSS more alike to $t\bar{t}$ JSS than QCD JSS
- High r_{inv} SVJs are more SM-like

Differences with previous applications

- Existing classification tasks[1][2] are quite different from this one:

Classification task	Computer science paper[1] / MNIST task	HEP paper[2] / task	SVJ search
Data	MNIST images	Jet images	1D array of JSS features
Data representation	32×32 in [0, 1]	40×40 in [0, 1]	9 features, not all bounded
Number of dimensions	1024	1600	9
Network architecture	2D CNN	2D CNN	DNN
Classification	1 MNIST class as OOD	QCD vs $t\bar{t}$ $t\bar{t}$ vs QCD QCD vs SVJ	$t\bar{t}$ vs SVJ QCD vs SVJ QCD + $t\bar{t}$ vs SVJ

- Development of this SVJ tagger done on $t\bar{t}$ sample
- In the next slides, presented AUCs are the average AUC over all signal model hypotheses

[1] arXiv:2105.05735
[2] arXiv:2206.14225

MCMC initialization:

- In theory, MCMC convergence independent on the initial point
- However, in practice with short chain, initialization is crucial

Several commonly used initialization algorithms of the MCMC:

- Contrastive Divergence[1] (CD)
- Persistent CD[2] (PCD)

CD[3]

- Initial distribution from training data
- Re-initialization after each parameter update (*i.e.* epoch)

PCD[4]

- Random initial distribution for first MCMC
- The model changes only slightly during parameter update
- Thus, for subsequent chains, initialize chain at the state in which it ended for the previous model
- Possibility to randomly re-initialize a small fraction of the samples

[1] Neural Comput 2002; 14 (8)
[2] PCD paper
[3] Illustration in backup slide 15
[4] Illustration in backup slide 16

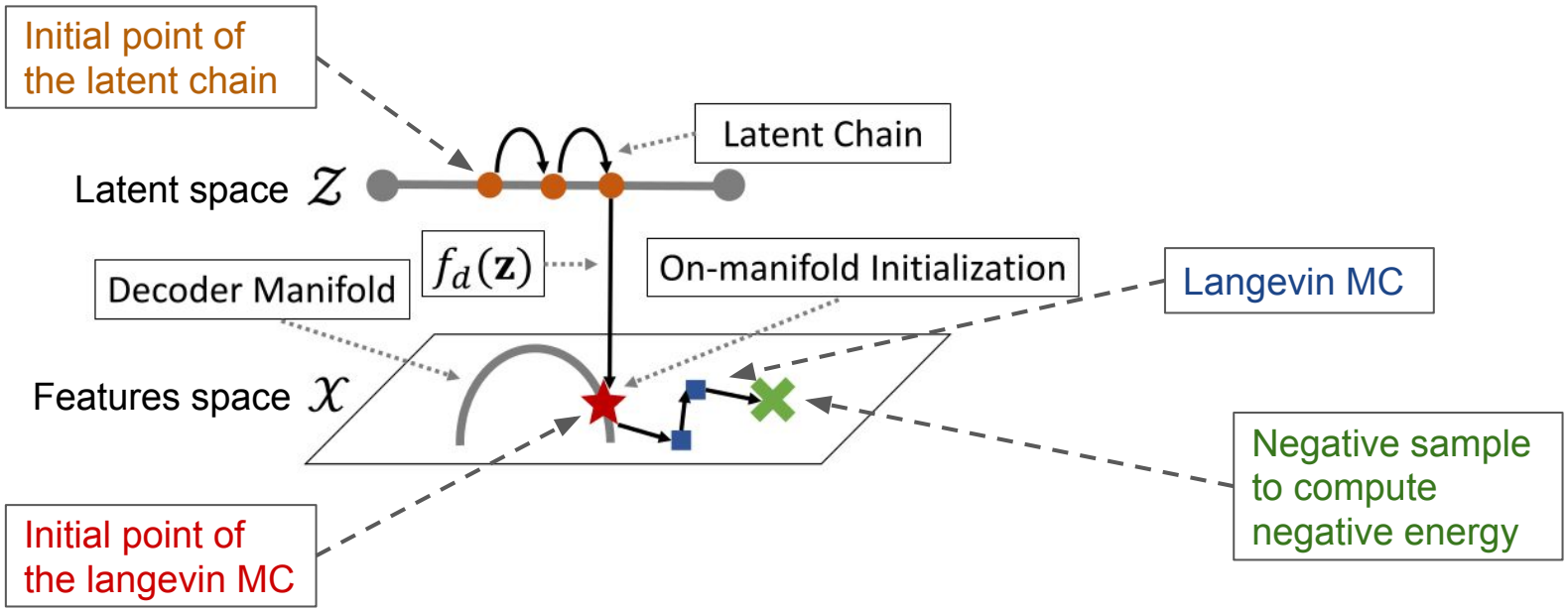
On-Manifold Initialization

Tailored MCMC initialization algorithm for AEs:

- CD and PCD have failure modes
  - CD failure mode: spurious low reconstruction error phase-space far from the training dataset
  - PCD failure mode: MCMC chains very correlated, spurious low reconstruction error phase-space can be missed

➔ Tailored algorithm for AE: On-Manifold Initialization (OMI)

- Run a first MCMC in the latent space to generate samples lying near the decoder manifold
- Use them as initial points for the usual MCMC



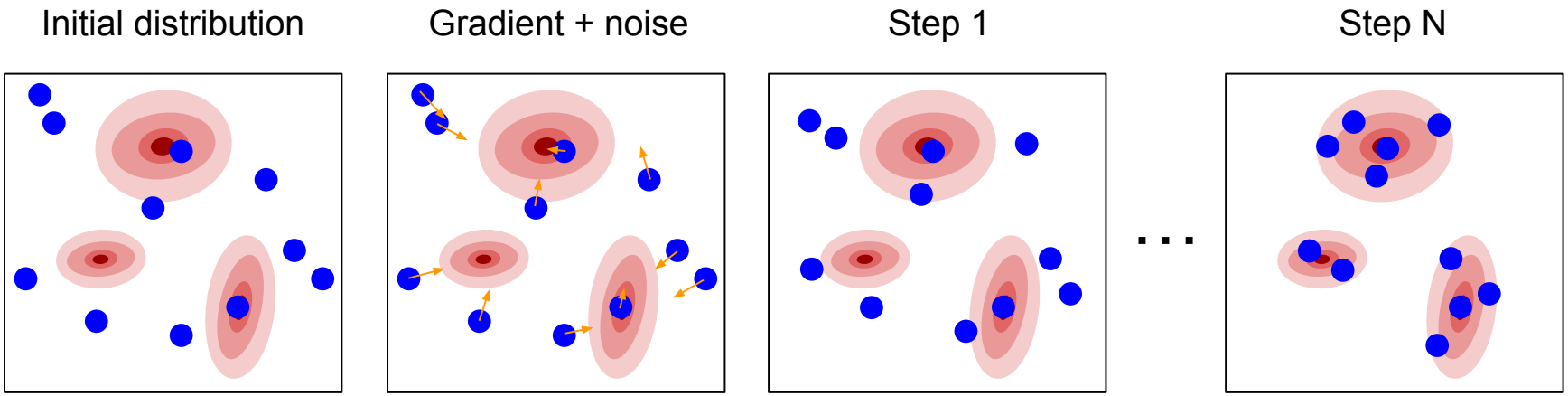
OMI versus (P)CD

- On image classification tasks, OMI was proved to outperform CD and PCD
- The performance severely depends on tuning the MCMC hyper-parameters

Table 1. MNIST hold-out class detection AUC scores. The values in parentheses denote the standard error of mean after 10 training runs.

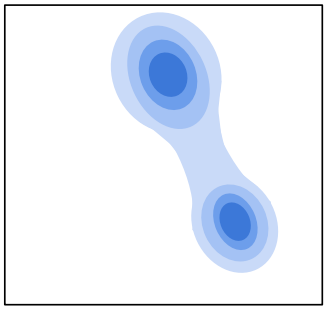
HOLD-OUT:	0	1	2	3	4	5	6	7	8	9	AVG
NAE-OMI	**.989**(.002)	**.919**(.013)	**.992**(.001)	**.949**(.004)	**.949**(.005)	**.978**(.003)	**.938**(.004)	**.975**(.024)	**.929**(.004)	**.934**(.005)	**.955**
NAE-CD	.799	.098	.878	.769	.656	.806	.874	.537	.876	.500	.679
NAE-PCD	.745	.114	.879	.754	.690	.813	.872	.509	.902	.544	.682
AE	.819	.131	.843	.734	.661	.755	.844	.542	.902	.537	.677
DAE	.769	.124	.872	.935	.884	.793	.865	.533	.910	.625	.731
VAE(R)	.954	.391	.978	.910	.860	.939	.916	.774	.946	.721	.839
VAE(L)	.967	.326	.976	.906	.798	.927	.928	.751	.935	.614	.813
WAE	.817	.145	.975	**.950**	.751	.942	.853	.912	.907	.799	.805
GLOW	.803	.014	.624	.625	.364	.561	.583	.326	.721	.426	.505
PXCNN++	.757	.030	.663	.663	.483	.642	.596	.307	.810	.497	.545
IGEBM	.926	.401	.642	.644	.664	.752	.851	.572	.747	.522	.672
DAGMM	.386	.304	.407	.435	.444	.429	.446	.349	.609	.420	.423

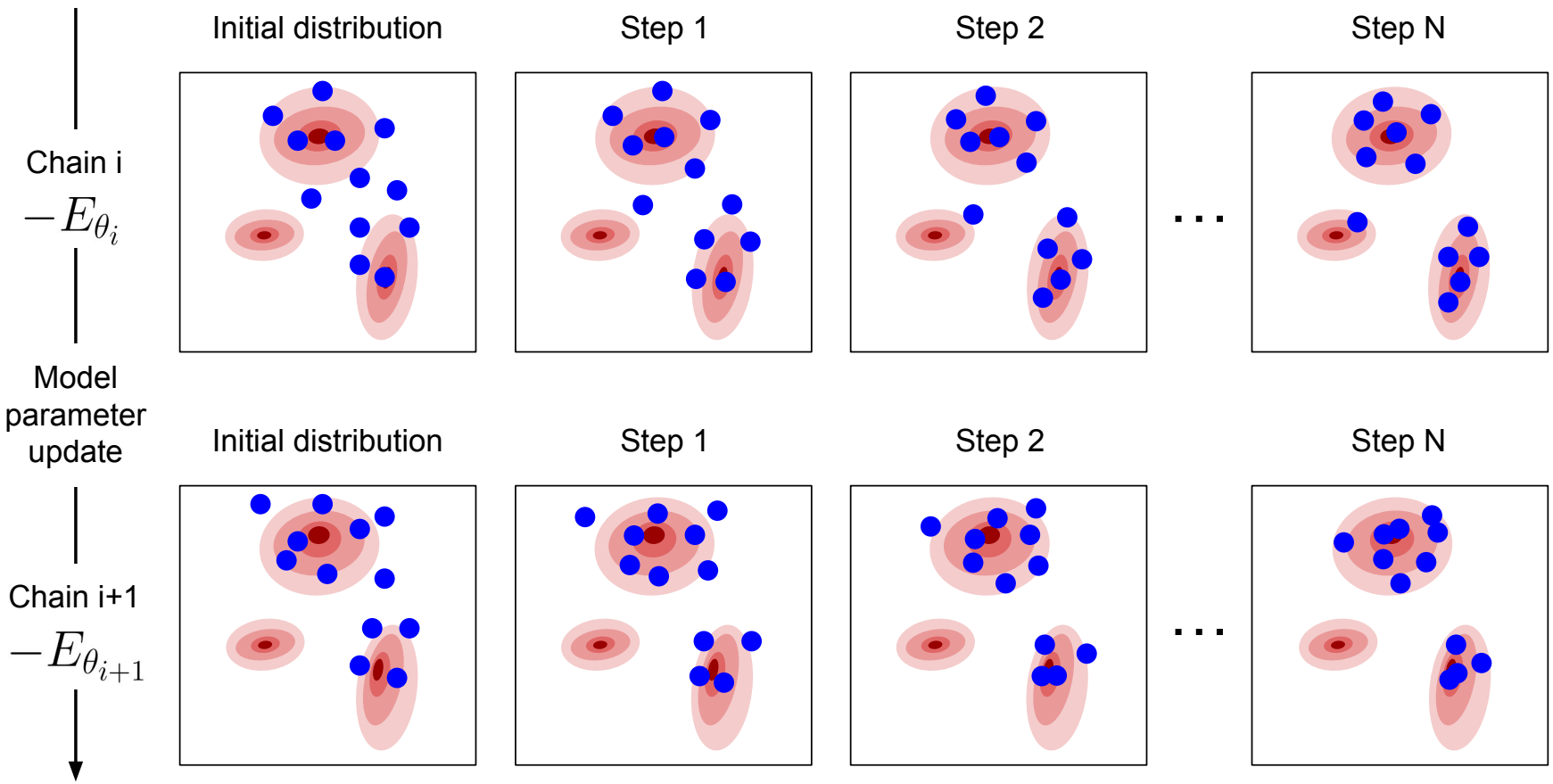
Principle of MCMC



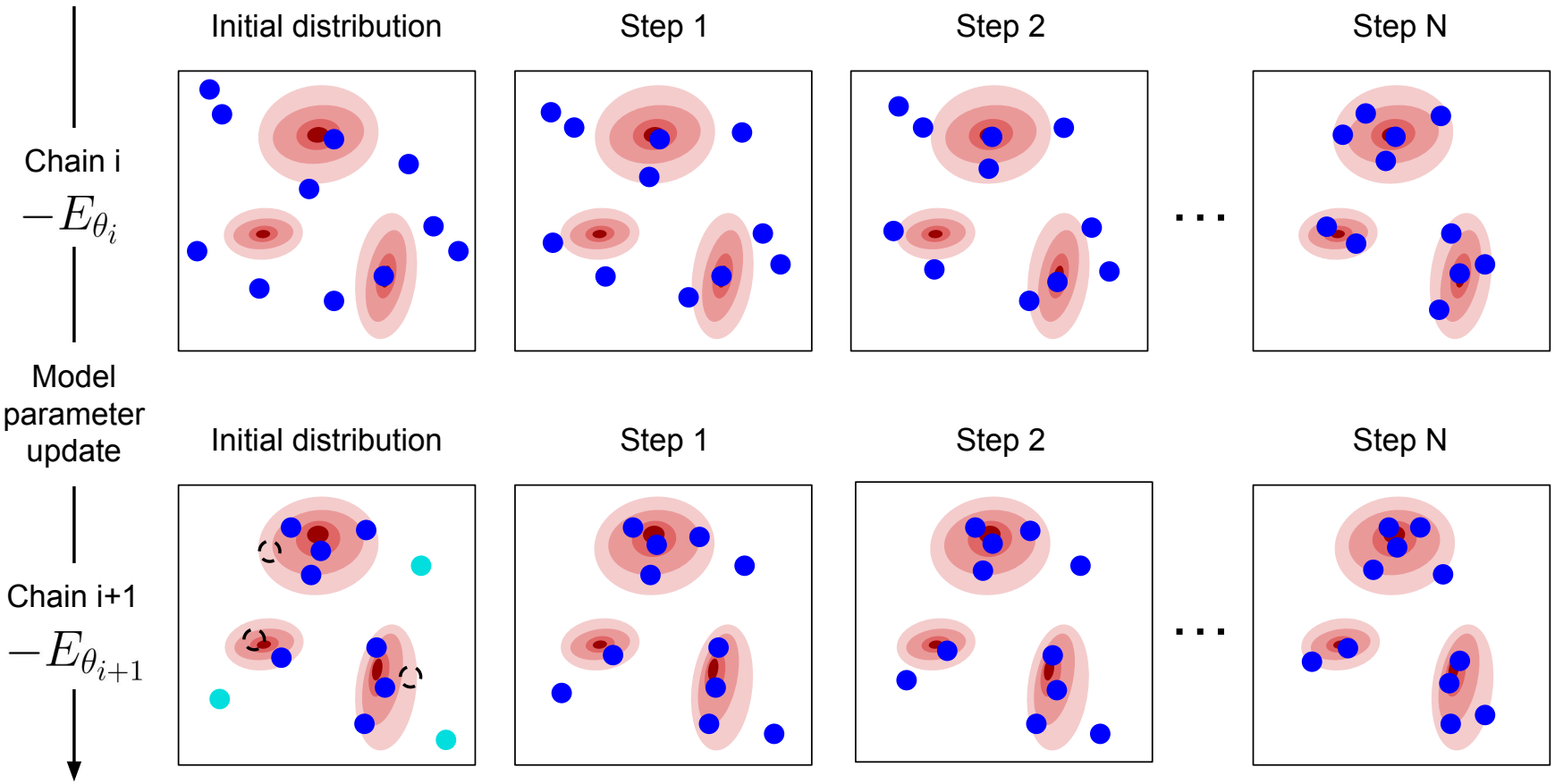
Example of a failure mode of CD: High probability mode far from training data distribution is not sampled

Training data distribution:





Principle of PCD

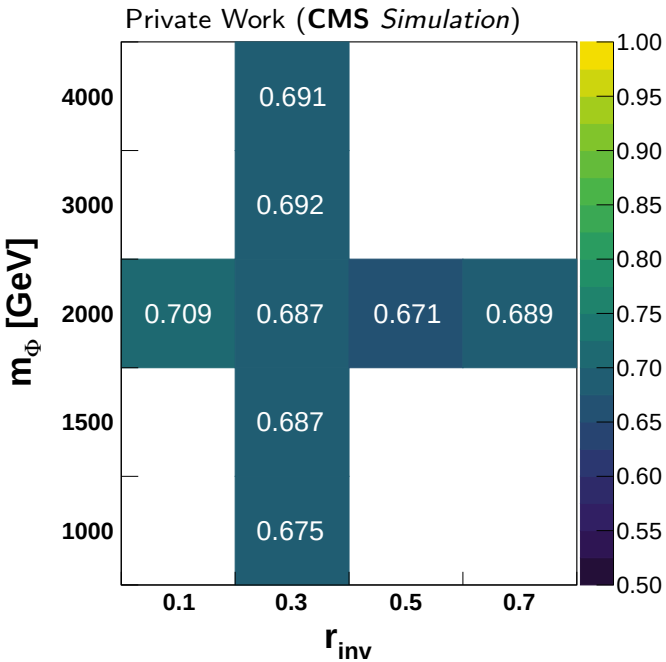


NAE against a mixture of $t\bar{t}$ and QCD

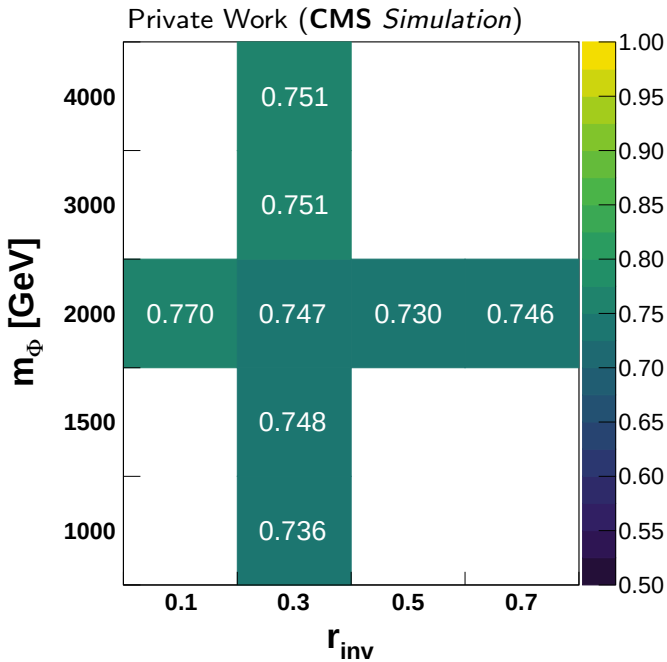
Disclaimer: "Cherry-picking" best model (at epoch with highest AUC)
NAE trained against a mixture of 50% QCD and 50% $t\bar{t}$

Successful unsupervised SVJ tagger against a mixture of $t\bar{t}$ and light-flavor jets!

$t\bar{t}$ vs SVJ



QCD vs SVJ

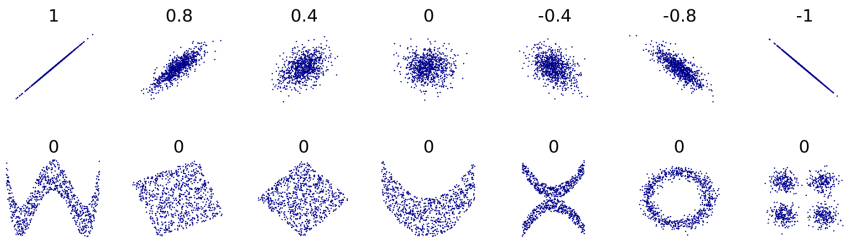


- Compared to separate background case, same performance for QCD, worse for $t\bar{t}$
- More important to better tag SVJ against QCD than $t\bar{t}$

Distance correlation (DisCo)

The **Pearson correlation** only evaluates **linear correlations**:

$$\rho^2_{\text{Pearson}}(X,Y) = \frac{\text{Cov}^2(X,Y)}{\text{Cov}(X,X)\text{Cov}(Y,Y)} \quad (9)$$



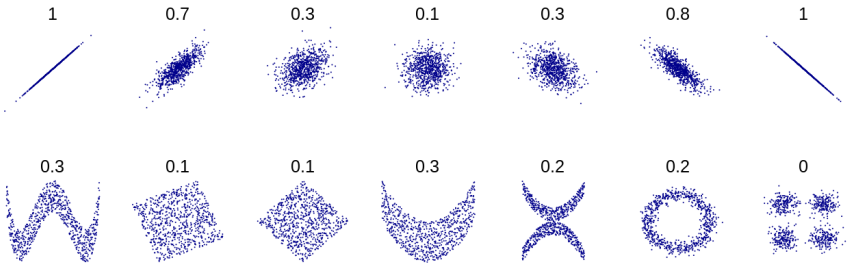
Pearson correlation coefficient

The **Distance correlation (DisCo)**[1] makes use of all information of the random variables:

$$\text{dCov}^2(X,Y) = \int d^p s d^q t \left| f_{X,Y}(s,t) - f_X(s) f_Y(t) \right|^2 w(s,t)$$

where f_X (resp. Y) is the characteristic function of X (resp. Y), $f_{X,Y}$ is the joint characteristic function of X and Y.
$f_{X,Y} == f_X f_Y$ iff X and Y are **independent**.

$$\text{DisCo}^2(X,Y) = \frac{\text{dCov}^2(X,Y)}{\text{dCov}(X,X)\text{dCov}(Y,Y)} \quad (10)$$



Distance correlation coefficient

[1] DisCo in ML for HEP at arXiv:2001.05310

ABCD method

Let V_1 and V_2 be 2 uncorrelated variables for the background distribution.

- Signal region A: $V_1 > c_1$ and $V_2 > c_2$
- Number of events in each region: N_A, N_B, N_C, N_D

Background estimation in signal region:

$$N_A^{\text{bkg}} = \frac{N_B N_C}{N_D}$$

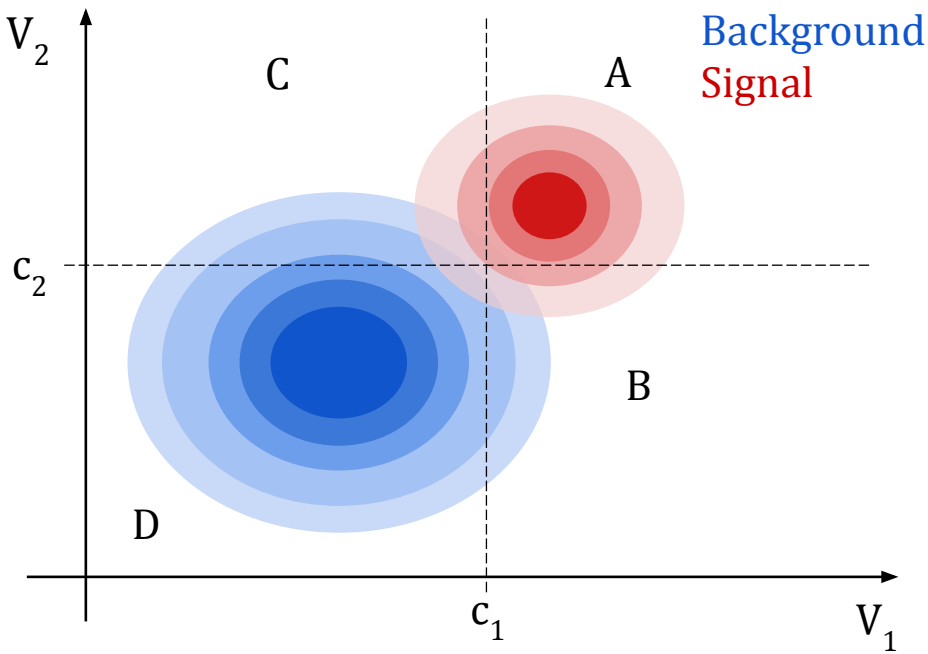


Illustration of the ABCD method