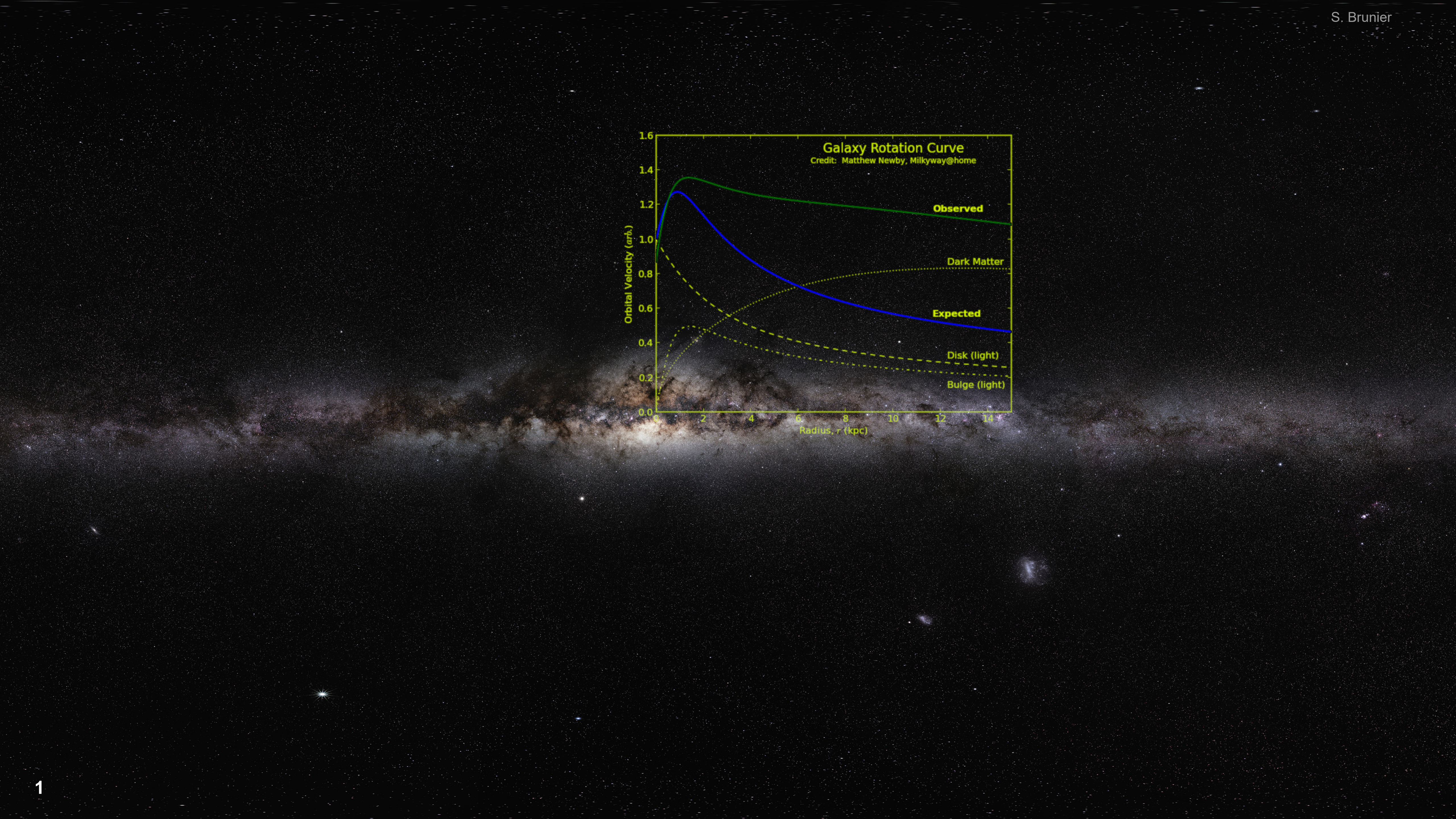
Karlsruhe Institute of Technology

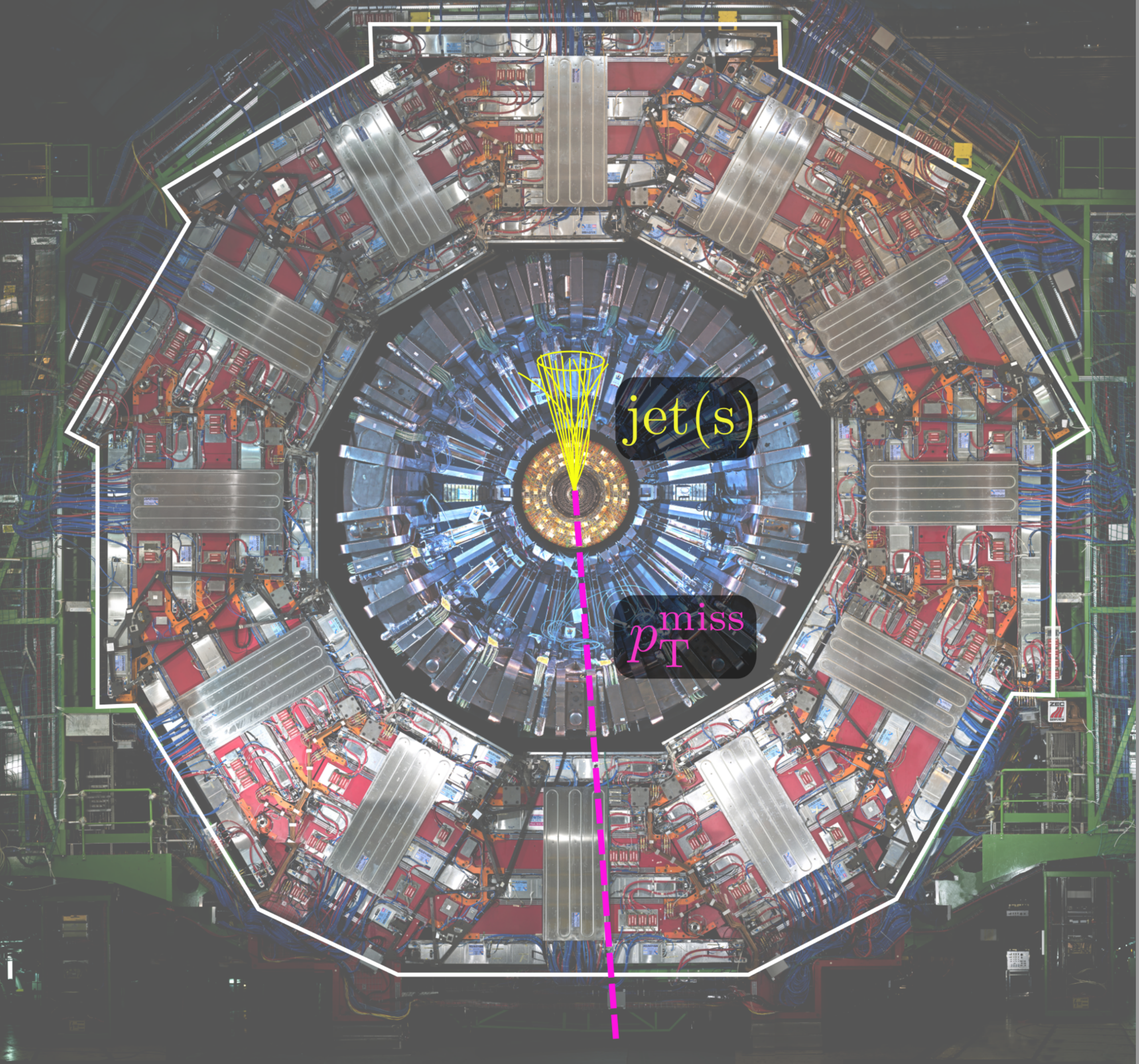
A Supervised Machine-Learning Search for Semi-Visible Jets in CMS Run-2 Scouting Data

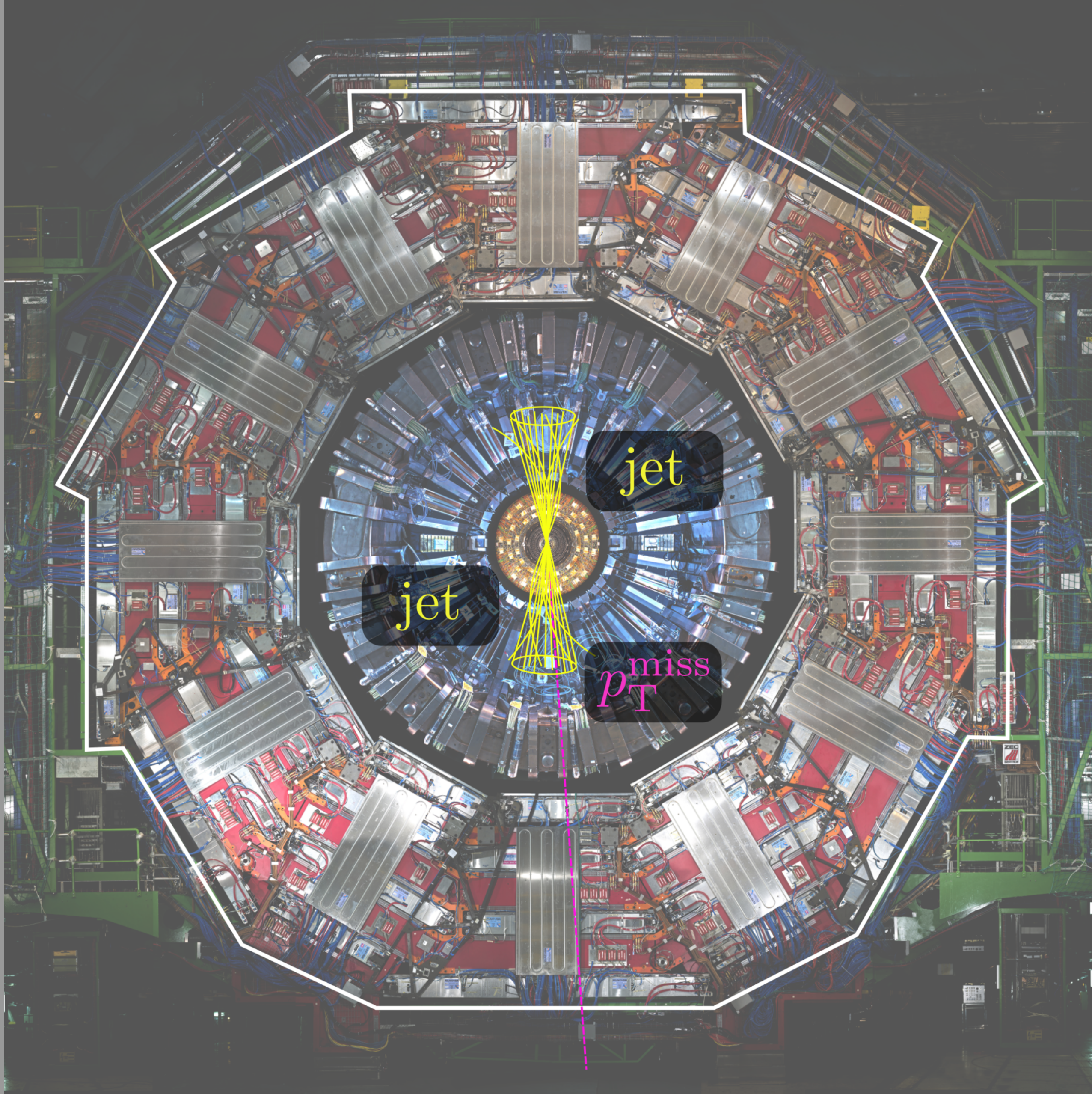
Jonas Janik · 12 Jan 2026


Galaxy Rotation Curve
Credit: Matthew Newby, Milkyway@home
Observed
Dark Matter
Expected
Disk (light)
Bulge (light)
Orbital Velocity (arb.)
Radius, r (kpc)
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
0
2
4
6
8
10
12
14

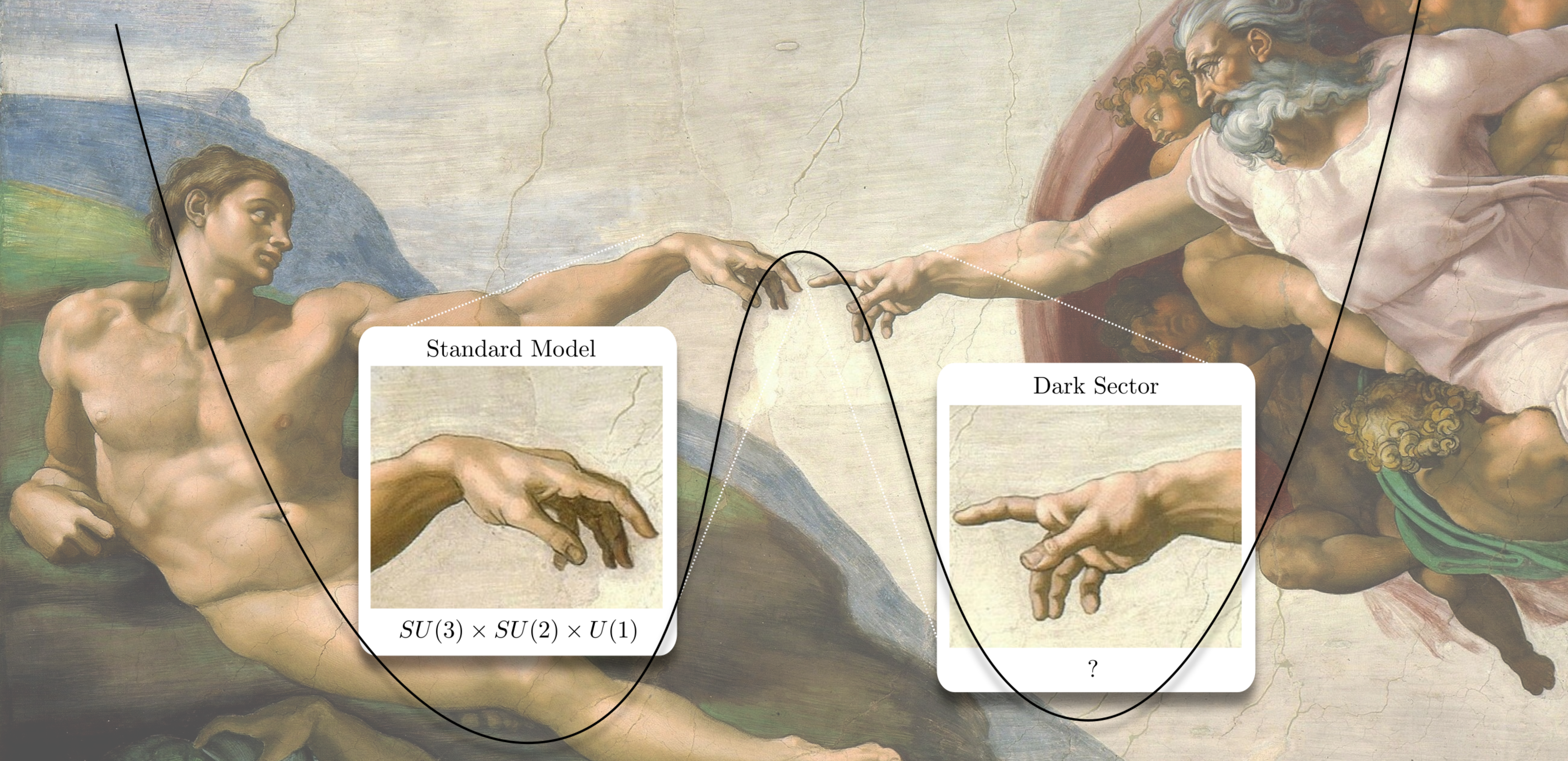
S. Brunier

back-to-back topology



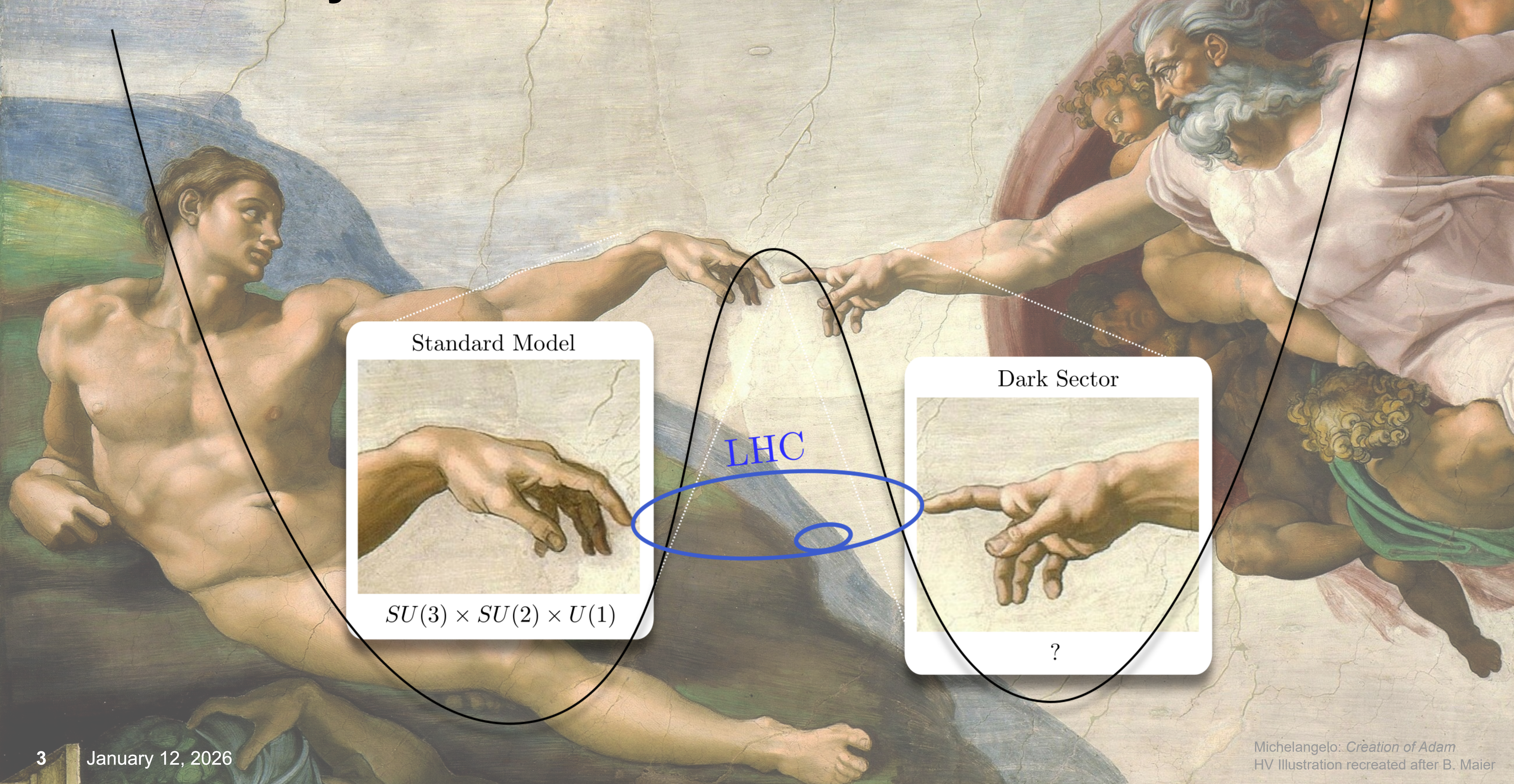


Hidden Valley Scenarios



Michelangelo: *Creation of Adam*
HV Illustration recreated after B. Maier

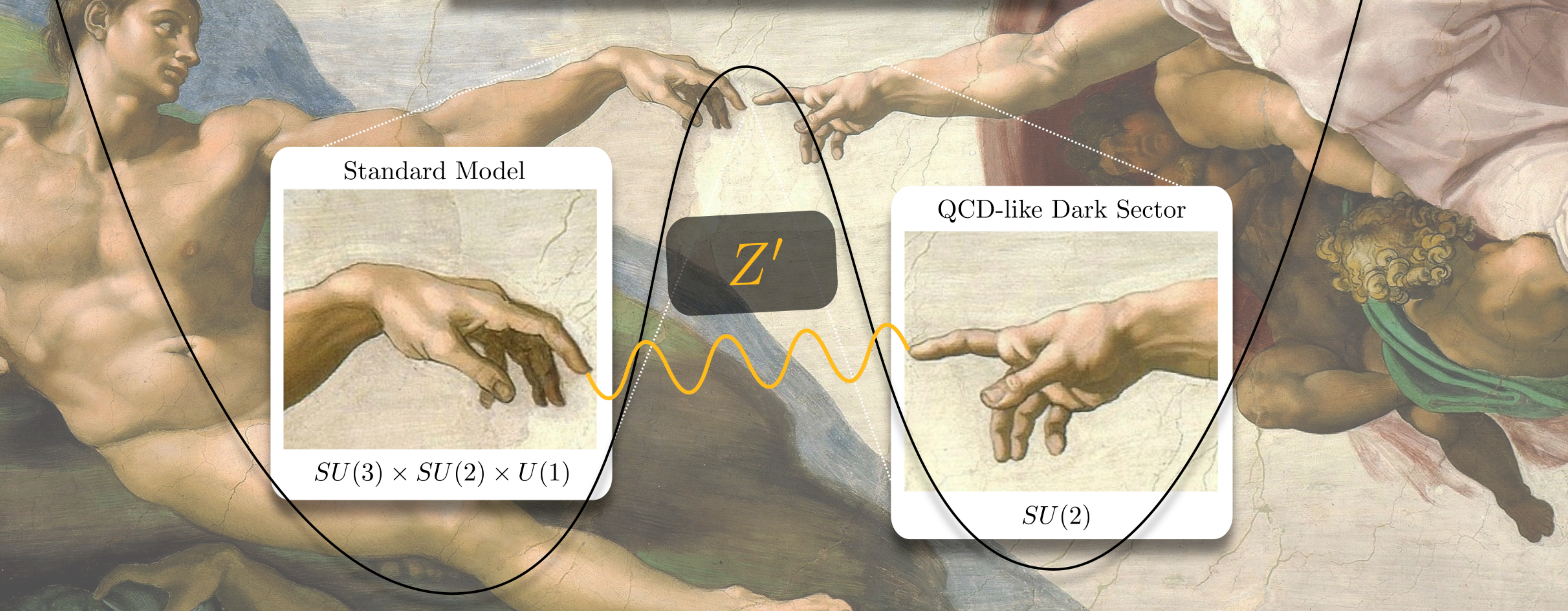
Hidden Valley Scenarios



Michelangelo: *Creation of Adam*
HV Illustration recreated after B. Maier

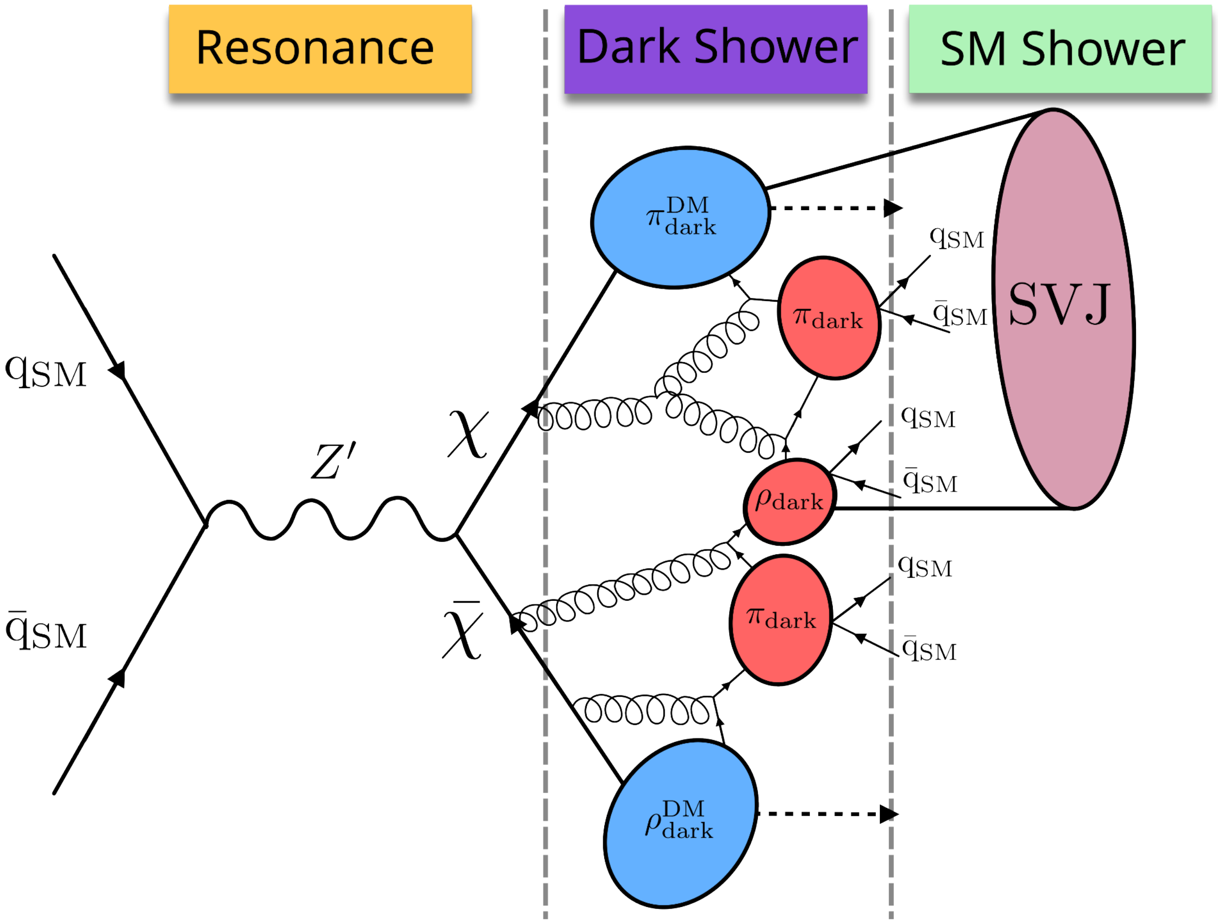
Hidden Valley Scenarios – A Benchmark Model

- Confining $SU(2)$ gauge group (2 dark colors, 2 dark flavors)
- 2 mass-degenerate dark quarks with $m_\chi = 10$ GeV
- Dark hadron mass $m_{\text{dark}} = 20$ GeV; prompt decays
- Leptophobic Z' mediator with $\text{BR}(Z' \to \chi\bar{\chi}) \simeq 1$

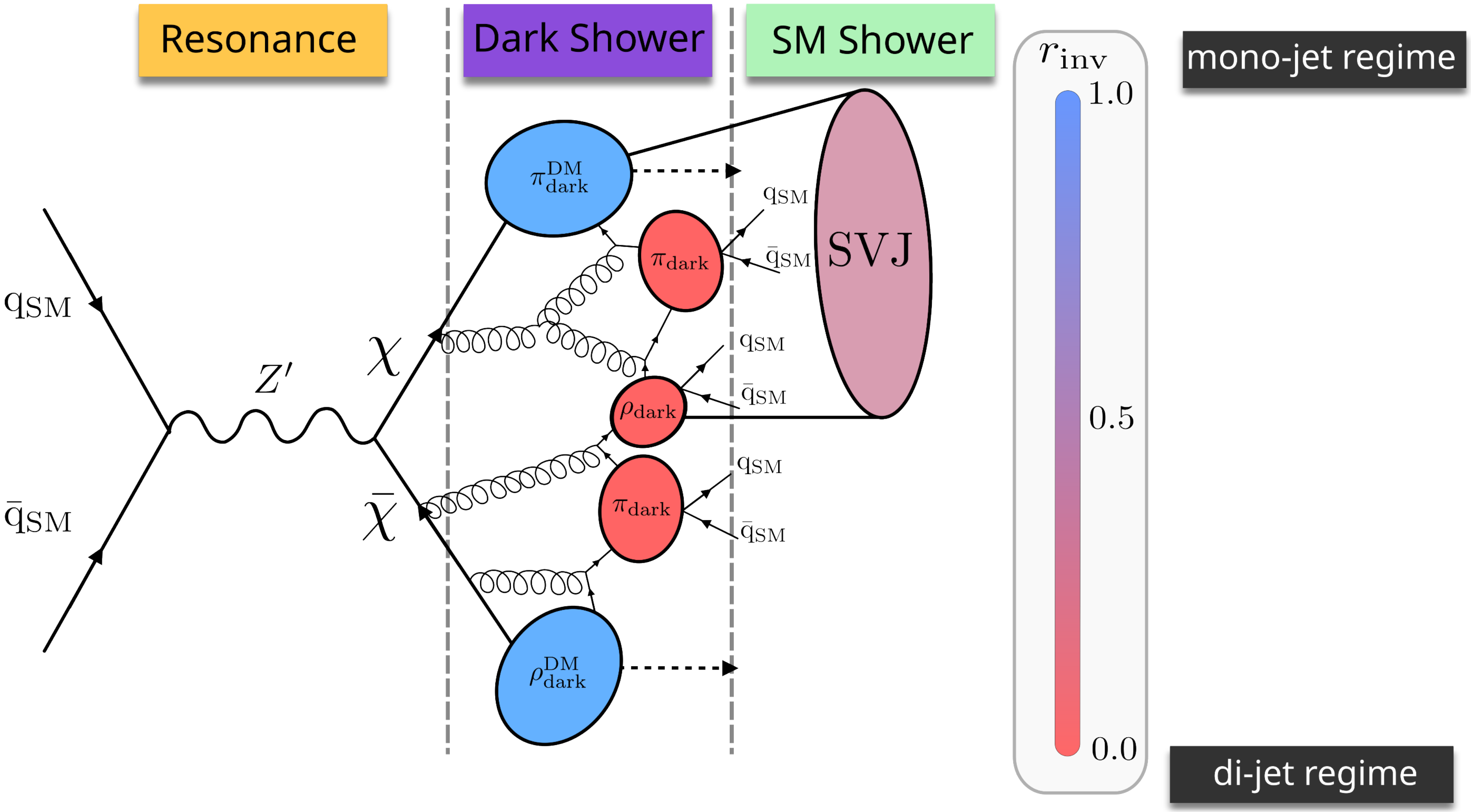


Michelangelo: *Creation of Adam*
HV Illustration recreated after B. Maier

From Hidden Valleys to Semi-Visible Jets

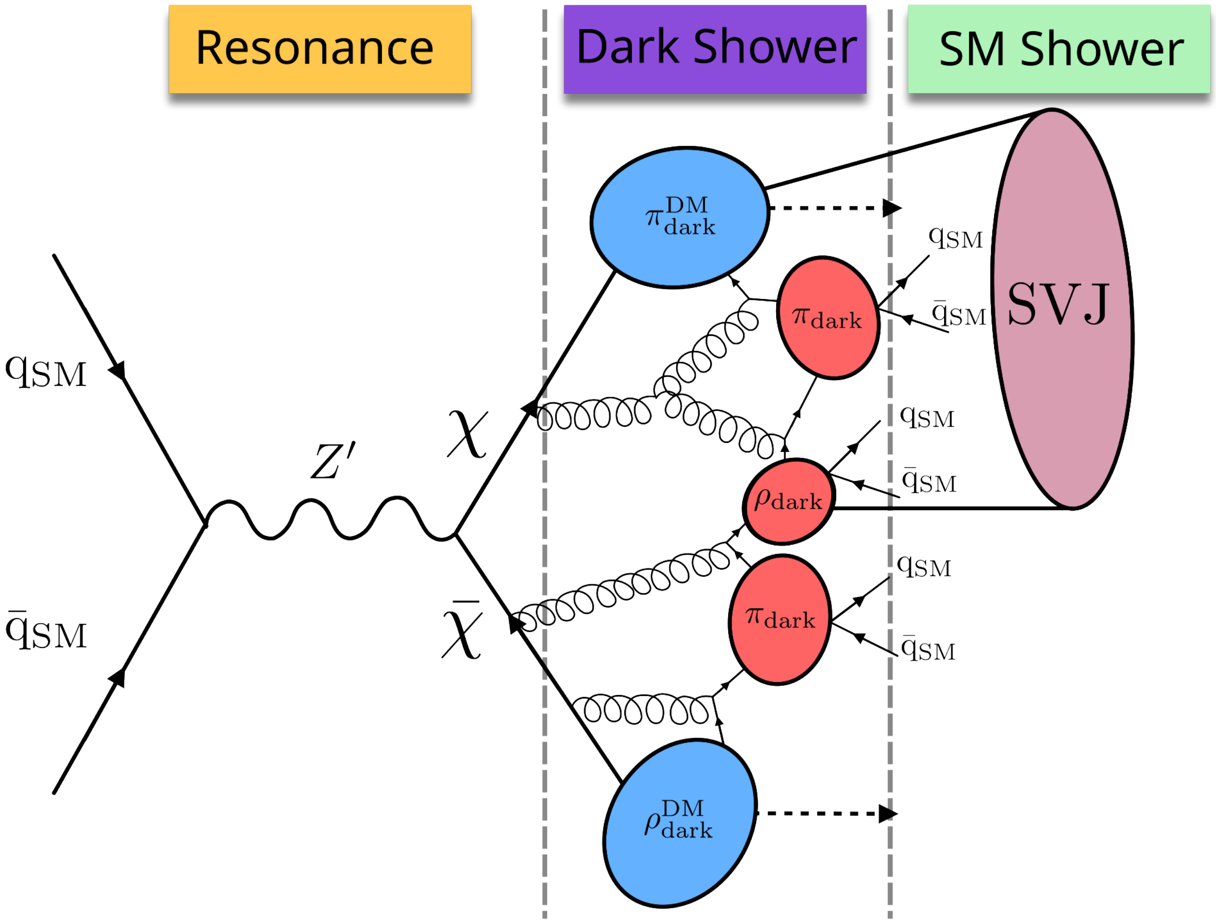


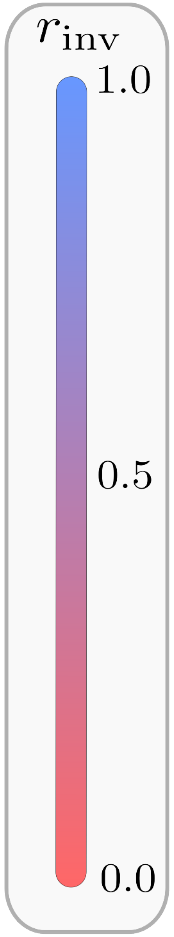
From Hidden Valleys to Semi-Visible Jets

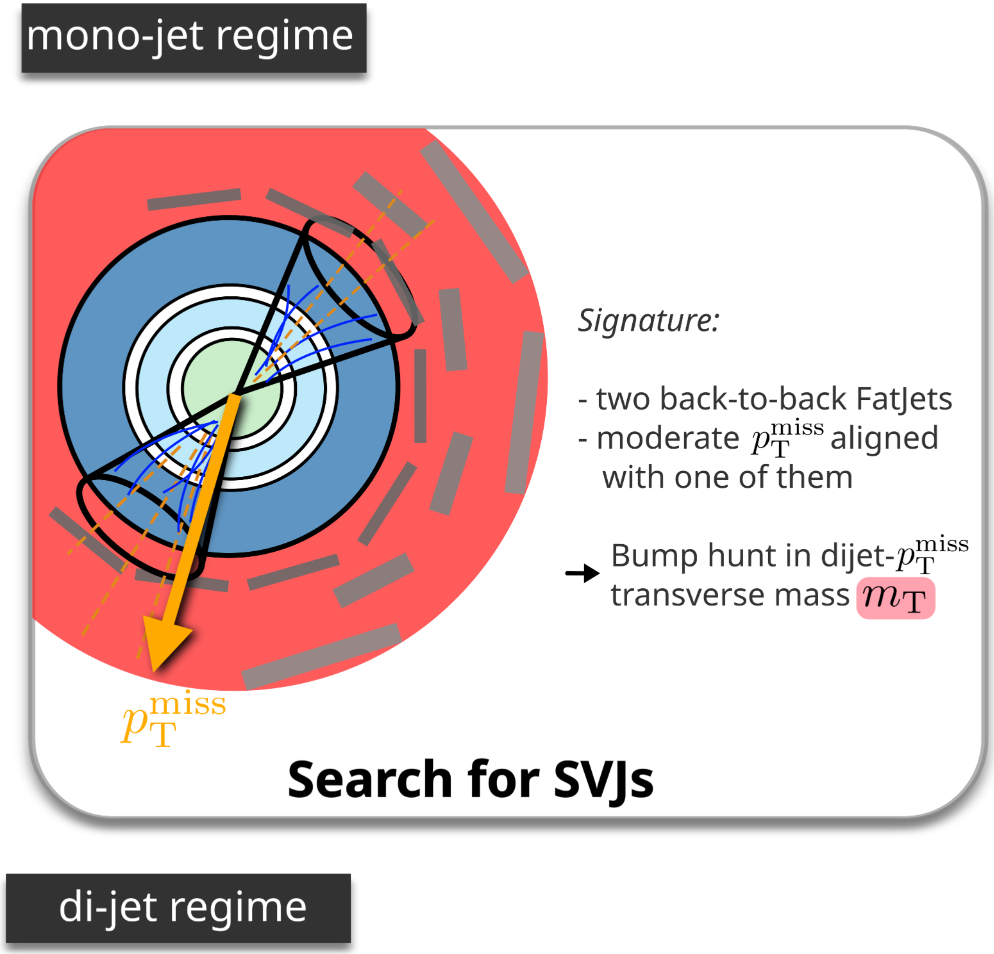


Scanned model parameters: $\boldsymbol{m_{Z'}}$, $\boldsymbol{r_{inv}}$, (m_{dark}), ...

From Hidden Valleys to Semi-Visible Jets







Signature:

- two back-to-back FatJets
- moderate p_T^{miss} aligned with one of them

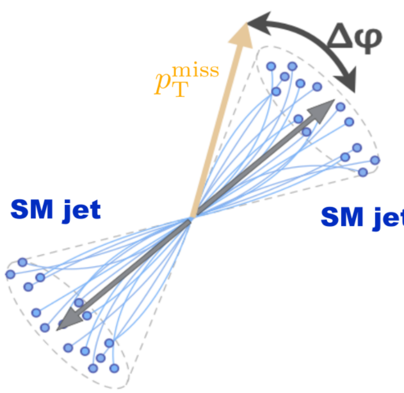
→ Bump hunt in dijet-p_T^{miss} transverse mass m_T

Scanned model parameters: $m_{Z'}$, r_{inv}, (m_{dark}), …

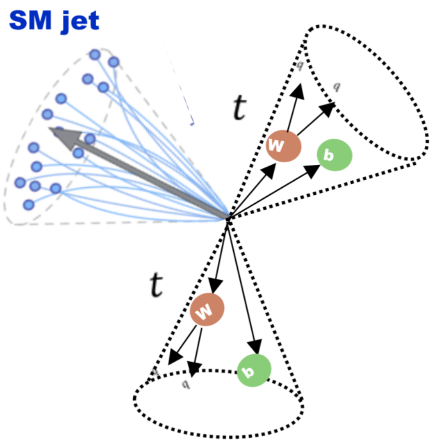
Standard Model Backgrounds

Karlsruhe Institute of Technology

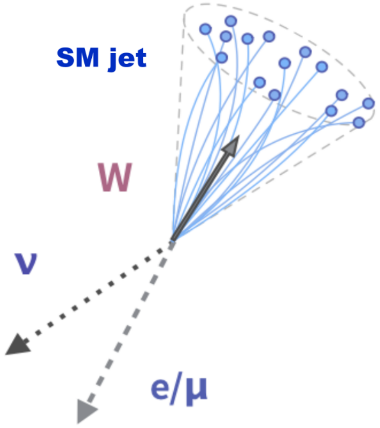
QCD multijet



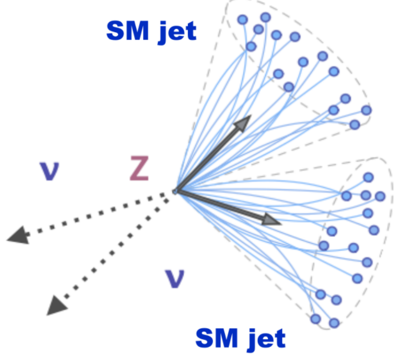
$t\bar{t}$



W+Jets



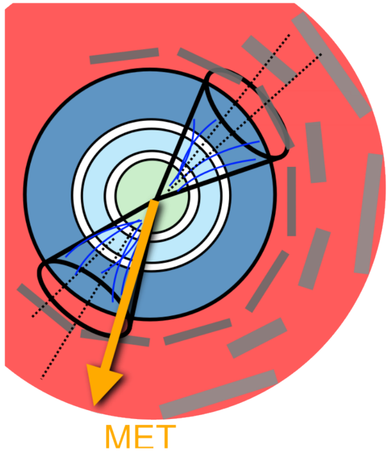
Z+Jets



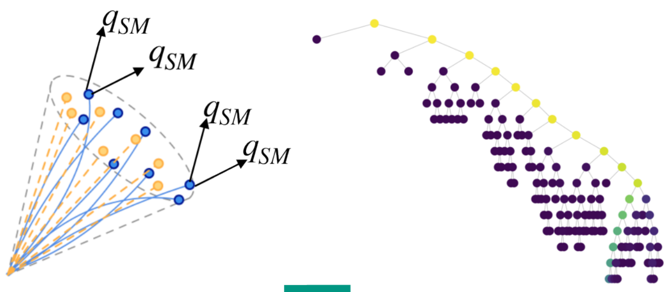
Adapted from https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-353690

Analysis Pipeline

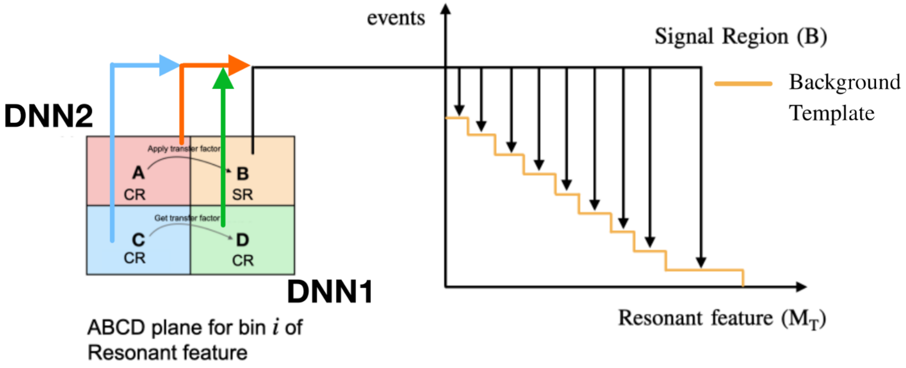
I) Trigger & Preselection



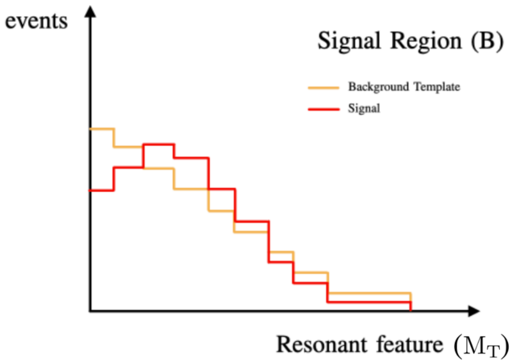
II) Supervised Jet Tagging (LundNET)



III) Event Tagging & ABCD (3D-ABCDisCoTEC)



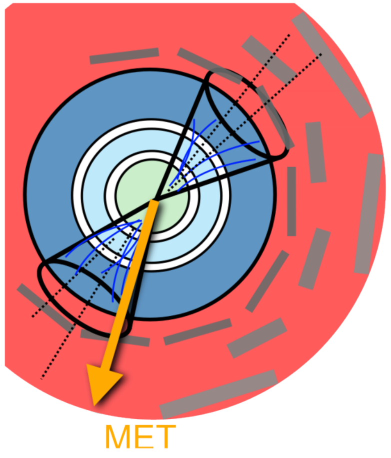
IV) Statistical Inference



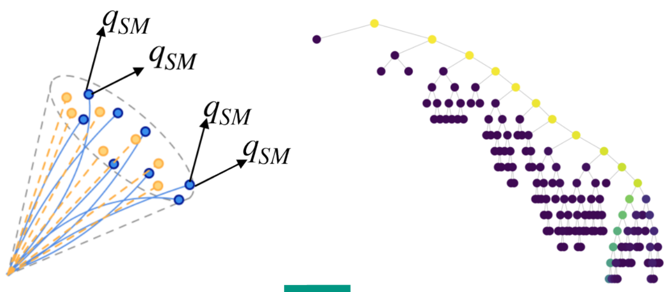
Adapted from https://cds.cern.ch/record/2940795

Analysis Pipeline

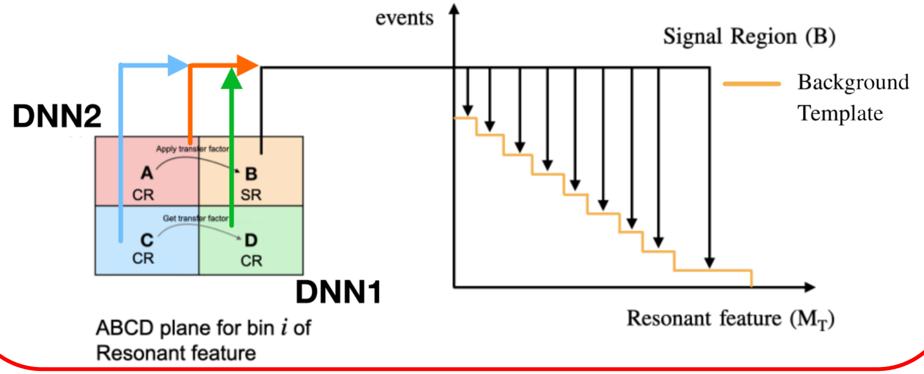
I) Trigger & Preselection



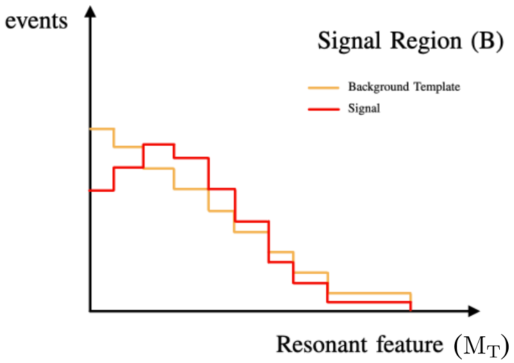
II) Supervised Jet Tagging (LundNET)



III) Event Tagging & ABCD (3D-ABCDisCoTEC)

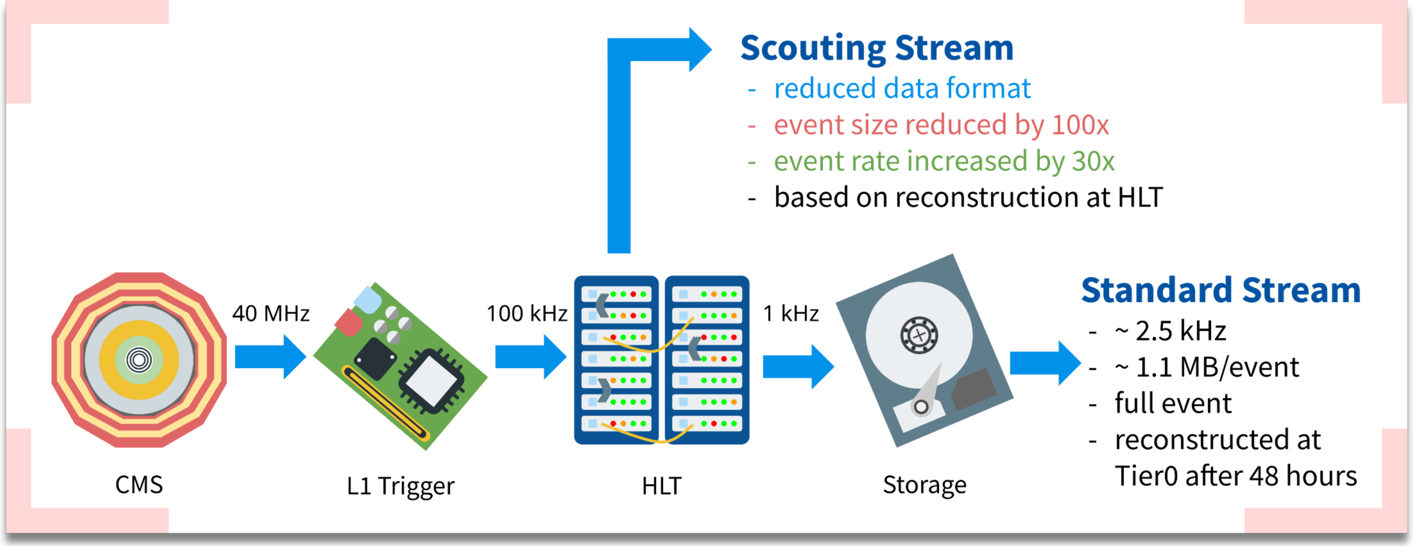


IV) Statistical Inference

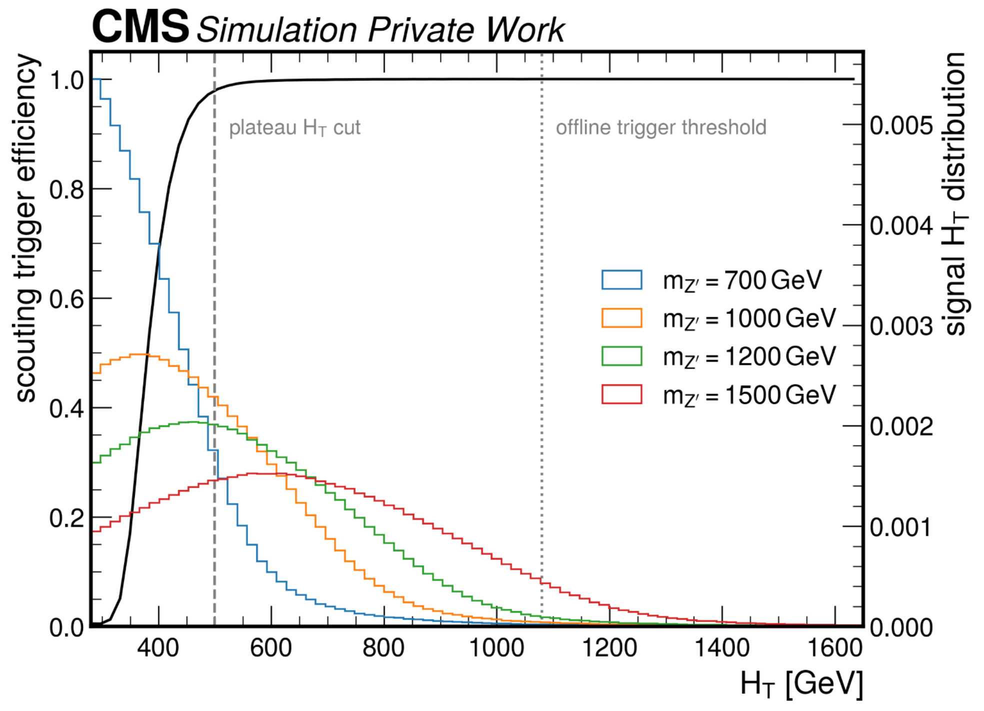


Adapted from https://cds.cern.ch/record/2940795

I) Lowering Trigger Thresholds: HLT-Scouting



Adapted from https://indico.cern.ch/event/1492840/contributions/6459931/



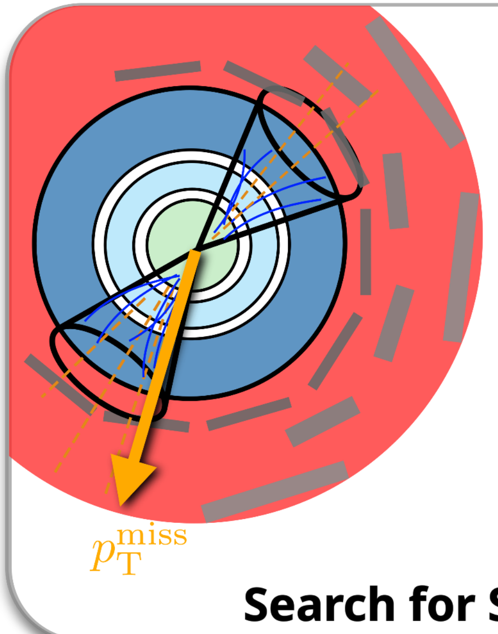
https://publish.etp.kit.edu/record/22315

- Offline trigger: $H_T^* > 1080$ GeV
- HLT-Scouting trigger: $H_T^* > 410$ GeV

* Scalar sum of all AK4 jet p_T (reconstructed from PF Candidates at HLT level)

I) Preselection

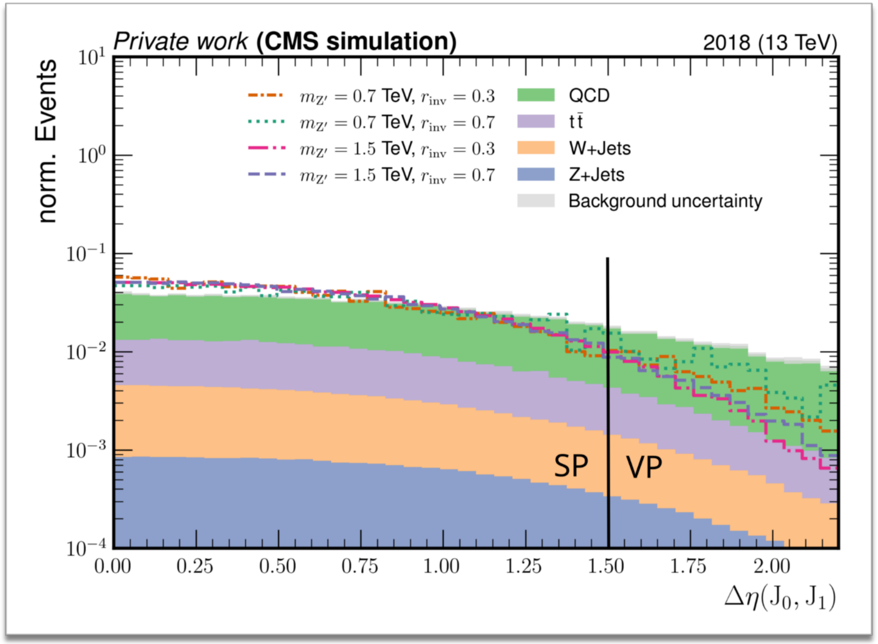
Selection criterion	Purpose
≥ 2 AK8 jets with $p_T \geq 150\,\text{GeV}$ and $\lvert\eta\rvert \leq 2.4$	signal topology
$M_T > 650\,\text{GeV}$	trigger efficiency
$R_T = p_T^{\text{miss}}/M_T > 0.24$	ensure missing p_T without sculpting M_T
$\Delta\phi_{\text{min}} = \min\left[\Delta\phi(J_0, p_T^{\text{miss}}),\ \Delta\phi(J_1, p_T^{\text{miss}})\right] < 0.8$	reject EW backgrounds
$\lvert\Delta\eta(J_0, J_1)\rvert < 2.2$	SP/VP definition



Signature:

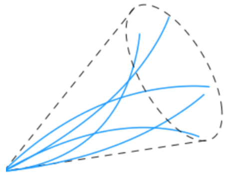
- two back-to-back FatJets
- moderate $p_{\text{T}}^{\text{miss}}$ aligned with one of them

→ Bump hunt in dijet-$p_{\text{T}}^{\text{miss}}$ invariant mass m_{T}

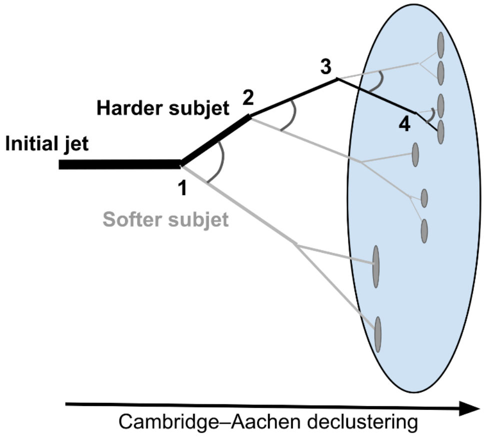


II) Supervised SVJ Tagging via LundNet

AK8 jet

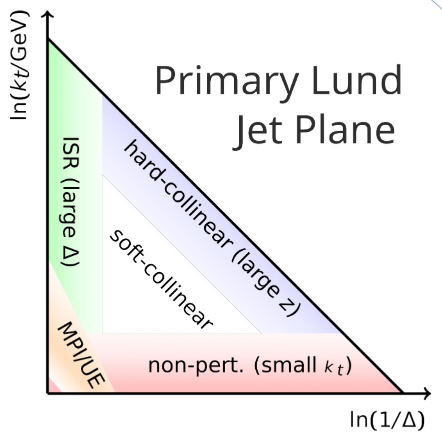


Recluster w/ C-A



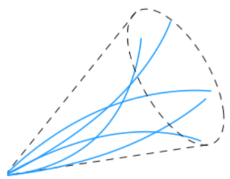
https://cds.cern.ch/record/2884797/

follow hardest branch only

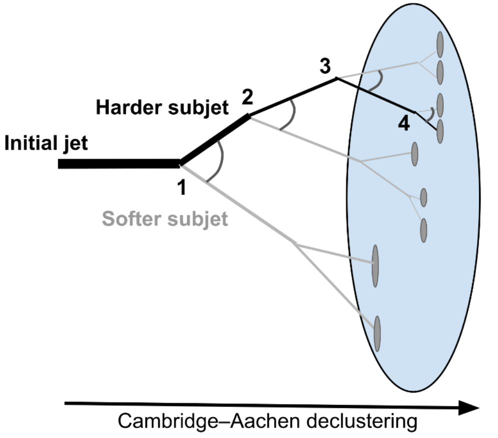


II) Supervised SVJ Tagging via LundNet

AK8 jet

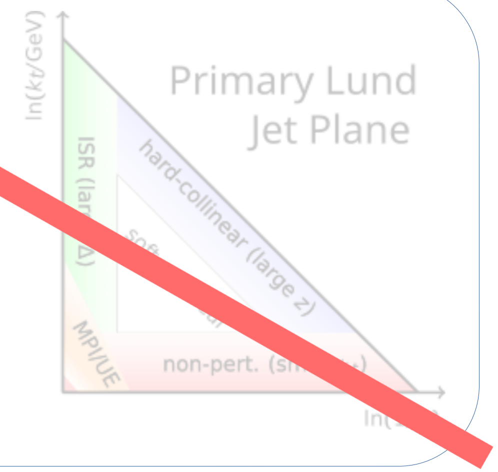


Recluster w/ C-A



https://cds.cern.ch/record/2884797/

follow hardest branch only



Lund Graph

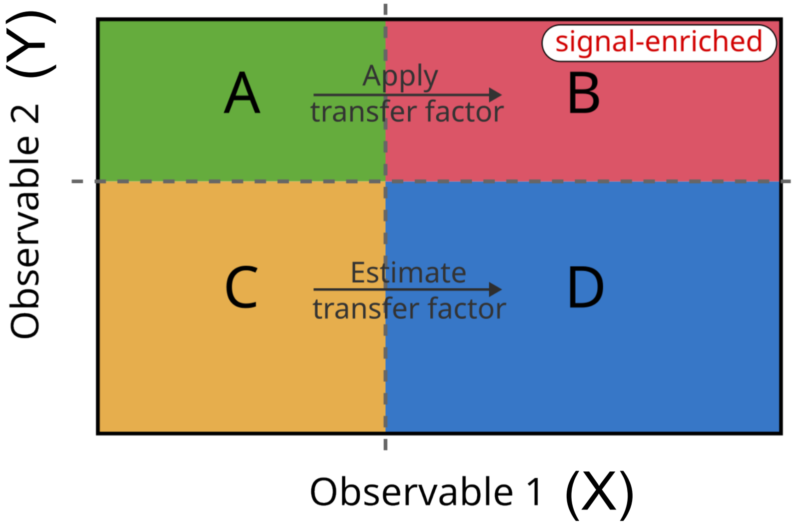
follow **all** branches, save tuple of kinematics at each split

GNN

LundScores (signal compatibility) for the two leading AK8 jets

III) The ABCD Strategy

$$\hat{n}_B^{\mathrm{b}} = n_A^{\mathrm{obs}} \times \frac{n_D^{\mathrm{obs}}}{n_C^{\mathrm{obs}}}$$

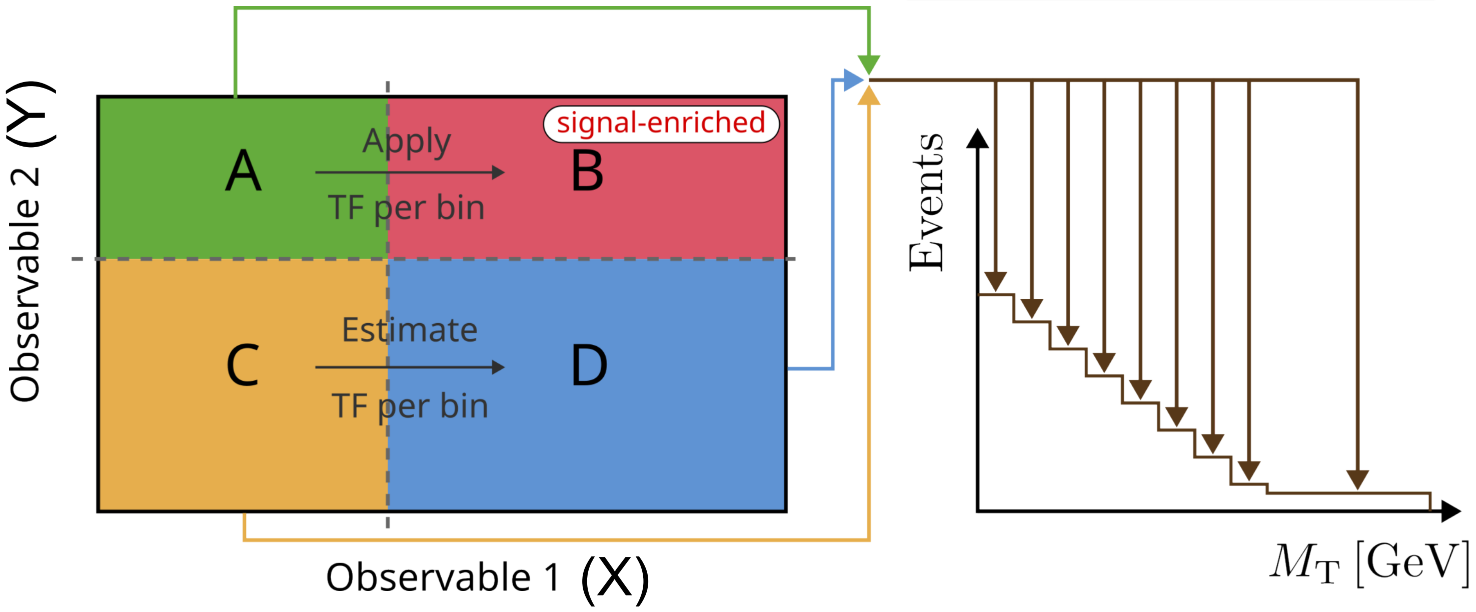


Closure conditions

- X, Y statistically independent
- signal contamination in CRs negligible

III) (Binned) ABCD Strategy

$$\hat{n}^{\mathrm{b}}_{B,i} = n^{\mathrm{obs}}_{A,i} \times \frac{n^{\mathrm{obs}}_{D,i}}{n^{\mathrm{obs}}_{C,i}}$$



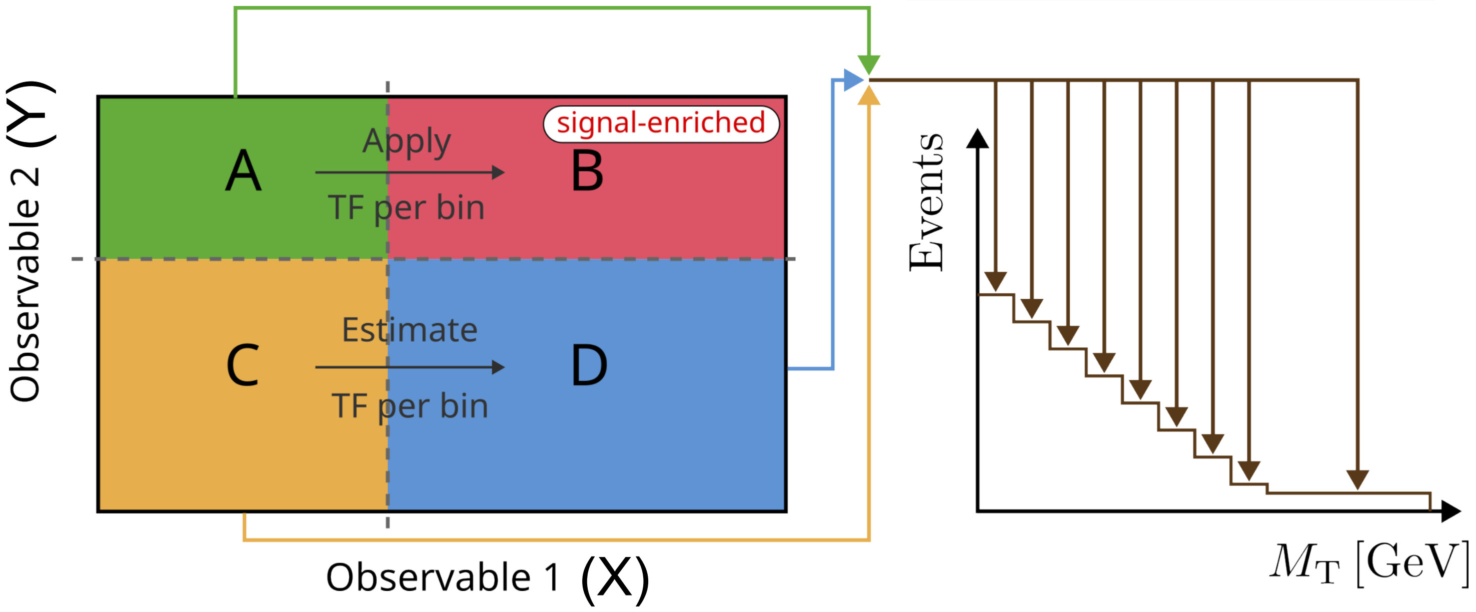
Closure conditions

- X, Y statistically independent
- signal contamination in CRs negligible
- X, Y uncorrelated to M_T

III) (Binned) ABCD Strategy

Karlsruhe Institute of Technology

$$\hat{n}^{\mathrm{b}}_{B,i} = n^{\mathrm{obs}}_{A,i} \times \frac{n^{\mathrm{obs}}_{D,i}}{n^{\mathrm{obs}}_{C,i}}$$



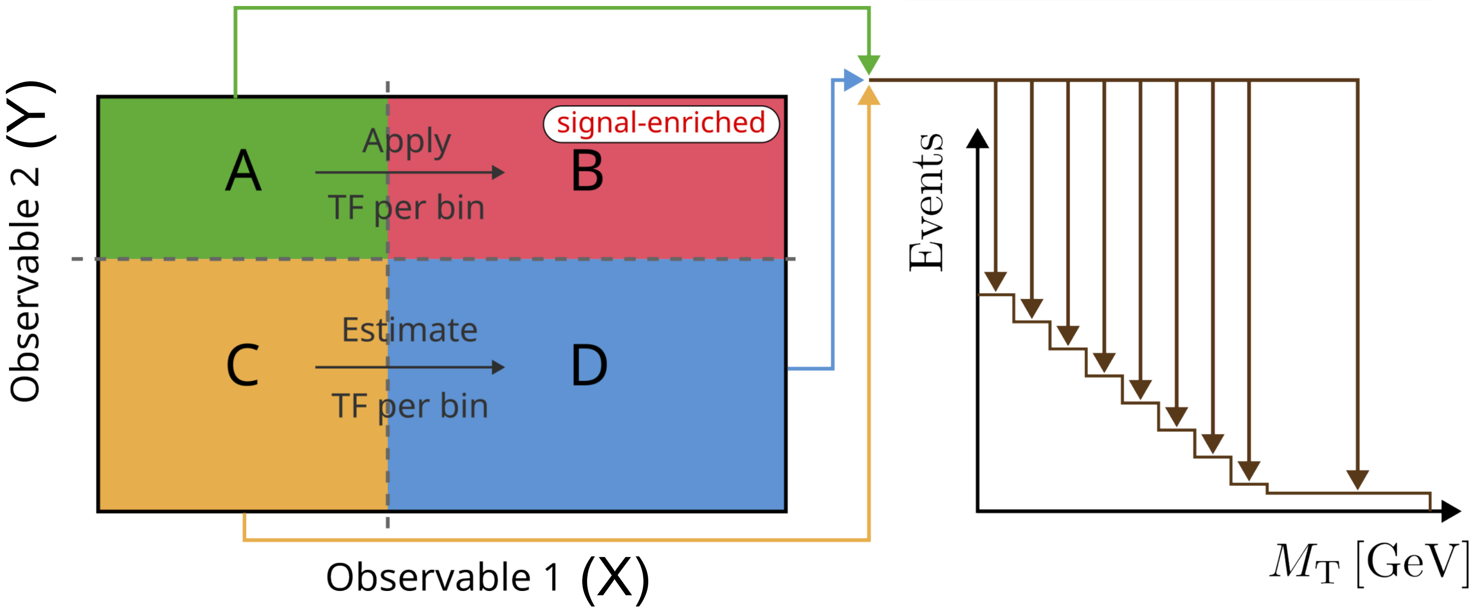
Closure conditions

- X, Y statistically independent
- signal contamination in CRs negligible
- X, Y uncorrelated to M_T

Problem: Find two Variables fulfilling all conditions

III) (Binned) ABCD Strategy

$$\hat{n}^{\mathrm{b}}_{B,i} = n^{\mathrm{obs}}_{A,i} \times \frac{n^{\mathrm{obs}}_{D,i}}{n^{\mathrm{obs}}_{C,i}}$$



Closure conditions

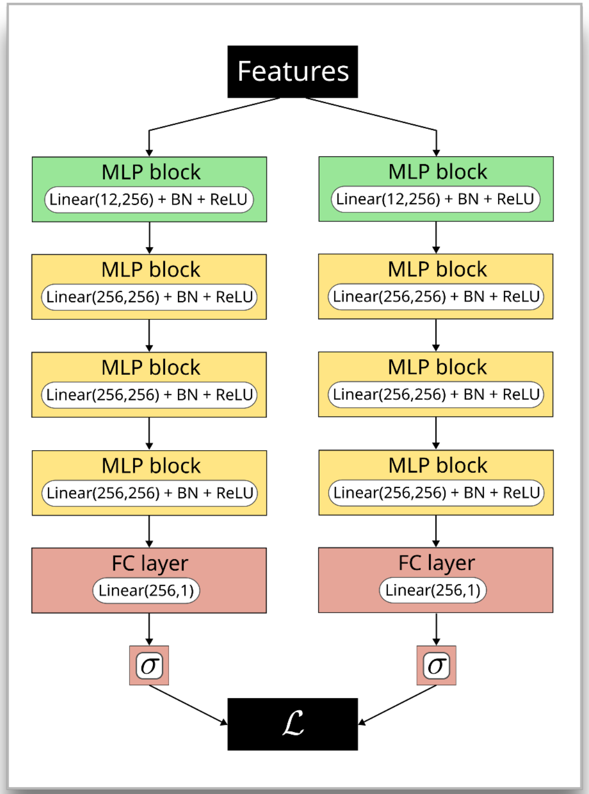
- X, Y statistically independent
- signal contamination in CRs negligible
- X, Y uncorrelated to M_T

Problem: Find two Variables fulfilling all conditions

Solution: Simultaneously train two DNNs, enforce conditions via loss terms

III) 3D-ABCDisCoTEC Method

- Apply global train/val/test split (40/20/40)
- **Train two MLPs simultaneously** (shared inputs, independent weights)
- Supervised training on MC mixture:
  QCD, ttbar, W+Jets, Z+Jets & signal grid (33 points)
- Event weights used; class-weight rescaling (S/B = 3/2)
- 5 Mini-batches (55,000 events) sampled per epoch with replacement



Input feature	Description
$p_{\mathrm{T}}^{\mathrm{miss}}$	missing transverse momentum
$\phi(\vec{p}_{\mathrm{T}}^{\mathrm{miss}})$	azimuthal angle of the missing transverse momentum
R_{T}	transverse ratio
$\Delta\phi_{\mathrm{min}}$	smaller azimuthal angle between $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ and each of the two leading jets
$\Delta\phi(\mathrm{J}_0, \mathrm{J}_1)$	$\Delta\phi$ between the leading jets
$\Delta\phi(\mathrm{J}_0, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})$	$\Delta\phi$ between the leading jet and $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$
$\Delta\phi(\mathrm{J}_1, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})$	$\Delta\phi$ between the subleading jet and $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$
$z(\mathrm{J}_0, \mathrm{J}_1)$	transverse momentum balance between the leading jets
$z(\mathrm{J}_0, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})$	transverse momentum balance between the leading jet and $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$
$z(\mathrm{J}_1, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})$	transverse momentum balance between the subleading jet and $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$
LundNet score (J_0)	LundNet jet tagger score of the leading jet
LundNet score (J_1)	LundNet jet tagger score of the subleading jet

$$z(\mathrm{J}_0, \mathrm{J}_1) = \frac{\min(p_{\mathrm{T}}(\mathrm{J}_0), p_{\mathrm{T}}(\mathrm{J}_1))}{p_{\mathrm{T}}(\mathrm{J}_0) + p_{\mathrm{T}}(\mathrm{J}_1)}$$

III) Constructing the Training Objective. . .

$$\mathcal{L}(\theta; \lambda) =$$

Closure conditions

- X, Y statistically independent
- signal contamination in CRs negligible
- X, Y uncorrelated to M_T

III) Constructing the Training Objective. . . .

$$\mathcal{L}(\theta;\lambda) = \underbrace{\lambda_{\text{BCE}}^{(X)}\,\mathcal{L}_{\text{BCE}}^{(X)} + \lambda_{\text{BCE}}^{(Y)}\,\mathcal{L}_{\text{BCE}}^{(Y)}}_{s-b\text{ discrimination}}$$

Closure conditions

- X, Y statistically independent
- signal contamination in CRs negligible
- X, Y uncorrelated to M_T

- BCE to maximize the significance in the signal region

III) Constructing the Training Objective. . . . .

$$\mathcal{L}(\theta;\lambda) = \underbrace{\lambda_{\text{BCE}}^{(X)}\,\mathcal{L}_{\text{BCE}}^{(X)} + \lambda_{\text{BCE}}^{(Y)}\,\mathcal{L}_{\text{BCE}}^{(Y)}}_{s-b\text{ discrimination}}$$

$$+\underbrace{\lambda_{\text{DisCo}}^{(X,Y)}\,\widehat{\text{dCorr}}_b(X,Y)}_{\text{decorrelation of } X \text{ and } Y}$$

Closure conditions

- X, Y statistically independent
- signal contamination in CRs negligible
- X, Y uncorrelated to M_T

- BCE to maximize the significance in the signal region

III) Constructing the Training Objective……

$$\mathcal{L}(\theta;\lambda) = \underbrace{\lambda_{\mathrm{BCE}}^{(X)}\,\mathcal{L}_{\mathrm{BCE}}^{(X)} + \lambda_{\mathrm{BCE}}^{(Y)}\,\mathcal{L}_{\mathrm{BCE}}^{(Y)}}_{s-b \text{ discrimination}}$$

$$+ \underbrace{\lambda_{\mathrm{DisCo}}^{(X,Y)}\,\widehat{\mathrm{dCorr}}_b(X,Y)}_{\text{decorrelation of } X \text{ and } Y}$$

$$+ \underbrace{\lambda_{\mathrm{DisCo}}^{(X,M_\mathrm{T})}\,\widehat{\mathrm{dCorr}}_b(X,M_\mathrm{T}) + \lambda_{\mathrm{DisCo}}^{(Y,M_\mathrm{T})}\,\widehat{\mathrm{dCorr}}_b(Y,M_\mathrm{T})}_{\text{decorrelation of } \{X,Y\} \text{ and } M_\mathrm{T}}$$

Closure conditions

- X, Y statistically independent
- signal contamination in CRs negligible
- X, Y uncorrelated to M_T

- BCE to maximize the significance in the signal region

III) Constructing the Training Objective. . . . . .

$$\mathcal{L}(\theta;\lambda) = \underbrace{\lambda_{\mathrm{BCE}}^{(X)}\,\mathcal{L}_{\mathrm{BCE}}^{(X)} + \lambda_{\mathrm{BCE}}^{(Y)}\,\mathcal{L}_{\mathrm{BCE}}^{(Y)}}_{s-b\text{ discrimination}}$$

$$+\underbrace{\lambda_{\mathrm{DisCo}}^{(X,Y)}\,\widehat{\mathrm{dCorr}}_b(X,Y)}_{\text{decorrelation of } X \text{ and } Y}$$

$$+\underbrace{\lambda_{\mathrm{DisCo}}^{(X,M_\mathrm{T})}\,\widehat{\mathrm{dCorr}}_b(X,M_\mathrm{T}) + \lambda_{\mathrm{DisCo}}^{(Y,M_\mathrm{T})}\,\widehat{\mathrm{dCorr}}_b(Y,M_\mathrm{T})}_{\text{decorrelation of } \{X,Y\} \text{ and } M_\mathrm{T}}$$

$$+\underbrace{\lambda_{\mathrm{DisCo}}^{(X,\Delta\eta)}\,\widehat{\mathrm{dCorr}}_b(X,\Delta\eta) + \lambda_{\mathrm{DisCo}}^{(Y,\Delta\eta)}\,\widehat{\mathrm{dCorr}}_b(Y,\Delta\eta)}_{\text{decorrelation of } \{X,Y\} \text{ and } \Delta\eta}$$

Closure conditions

- X, Y statistically independent
- signal contamination in CRs negligible
- X, Y uncorrelated to M_T

- BCE to maximize the significance in the signal region

- decorrelate X, Y against $\Delta\eta$ to encourage closure in signal and validation plane

III) Constructing the Training Objective……..

Karlsruhe Institute of Technology

$$\mathcal{L}(\theta;\lambda) = \underbrace{\lambda^{(X)}_{\mathrm{BCE}}\,\mathcal{L}^{(X)}_{\mathrm{BCE}} + \lambda^{(Y)}_{\mathrm{BCE}}\,\mathcal{L}^{(Y)}_{\mathrm{BCE}}}_{s-b\ \text{discrimination}}$$

$$+ \underbrace{\lambda^{(X,Y)}_{\mathrm{DisCo}}\,\widehat{\mathrm{dCorr}}_b(X,Y)}_{\text{decorrelation of } X \text{ and } Y}$$

$$+ \underbrace{\lambda^{(X,M_\mathrm{T})}_{\mathrm{DisCo}}\,\widehat{\mathrm{dCorr}}_b(X,M_\mathrm{T}) + \lambda^{(Y,M_\mathrm{T})}_{\mathrm{DisCo}}\,\widehat{\mathrm{dCorr}}_b(Y,M_\mathrm{T})}_{\text{decorrelation of } \{X,Y\} \text{ and } M_\mathrm{T}}$$

$$+ \underbrace{\lambda^{(X,\Delta\eta)}_{\mathrm{DisCo}}\,\widehat{\mathrm{dCorr}}_b(X,\Delta\eta) + \lambda^{(Y,\Delta\eta)}_{\mathrm{DisCo}}\,\widehat{\mathrm{dCorr}}_b(Y,\Delta\eta)}_{\text{decorrelation of } \{X,Y\} \text{ and } \Delta\eta}$$

$$+ \underbrace{\sum_{k\in\{\mathrm{SP},\mathrm{VP}\}} \lambda^{(k)}_{\not{C}}\,\not{C}^{b}_{k}}_{\text{normalization closure}}$$

$$+ \underbrace{\sum_{k\in\{\mathrm{SP},\mathrm{VP}\}} \sum_{i=1}^{N_\mathrm{bins}} \lambda^{(k)}_{\not{C},i}\,\not{C}^{b}_{k,i}}_{\text{shape-level closure}}\,.$$

Closure conditions

- X, Y statistically independent
- signal contamination in CRs negligible
- X, Y uncorrelated to M_T

- BCE to maximize the significance in the signal region
- decorrelate X, Y against $\Delta\eta$ to encourage closure in signal and validation plane
- enhance closure in normalization [1] and shape (novelty)

[1] https://arxiv.org/pdf/2506.08826

III) Constructing the Training Objective

$$\mathcal{L}(\theta;\lambda) = \underbrace{\lambda^{(X)}_{\mathrm{BCE}}\,\mathcal{L}^{(X)}_{\mathrm{BCE}} + \lambda^{(Y)}_{\mathrm{BCE}}\,\mathcal{L}^{(Y)}_{\mathrm{BCE}}}_{s-b\text{ discrimination}}$$

$$+ \underbrace{\lambda^{(X,Y)}_{\mathrm{DisCo}}\,\widehat{\mathrm{dCorr}}_b(X,Y)}_{\text{decorrelation of } X \text{ and } Y}$$

$$+ \underbrace{\lambda^{(X,M_{\mathrm{T}})}_{\mathrm{DisCo}}\,\widehat{\mathrm{dCorr}}_b(X,M_{\mathrm{T}}) + \lambda^{(Y,}_{\mathrm{Dis}}}_{\text{decorrelation of } \{X,Y\}}$$

$$+ \underbrace{\lambda^{(X,\Delta\eta)}_{\mathrm{DisCo}}\,\widehat{\mathrm{dCorr}}_b(X,\Delta\eta) + \lambda^{(Y,\Delta}_{\mathrm{DisC}}}_{\text{decorrelation of } \{X,Y\} \text{ a}}$$

$$+ \underbrace{\sum_{k\in\{\mathrm{SP,VP}\}} \lambda^{(k)}_{\mathcal{C}}\,\mathcal{C}^b_k}_{\text{normalization closure}}$$

$$+ \underbrace{\sum_{k\in\{\mathrm{SP,VP}\}} \sum_{i=1}^{N_{\mathrm{bins}}} \lambda^{(k)}_{\mathcal{C},i}\,\mathcal{C}^b_{k,i}}_{\text{shape-level closure}} \,.$$

Modified Differential Multiplier Method (MDMM)

- keep BCE sum as primary optimization target
- treat other loss terms as inequality constraints $\mathcal{L}_j(\theta) \leq \epsilon_j$

$$\phi_j(\theta, s_j) = \epsilon_j - \mathcal{L}_j(\theta) - s_j^2\,,$$

- Augmented Lagrangian formulation:

$$\tilde{\mathcal{L}}(\theta,\lambda,s) = \mathcal{L}^{(X)}_{\mathrm{BCE}} + \mathcal{L}^{(Y)}_{\mathrm{BCE}} + \sum_{j\in\mathcal{J}} \alpha_j \left(-\tilde{\lambda}_j\,\phi_j(\theta,s_j) + \tfrac{\rho_j}{2}\,\phi_j(\theta,s_j)^2\right),$$

- previously fixed loss weights λ become trainable multipliers
- update network parameters by gradient descent, multipliers by dual gradient ascent

Closure conditions

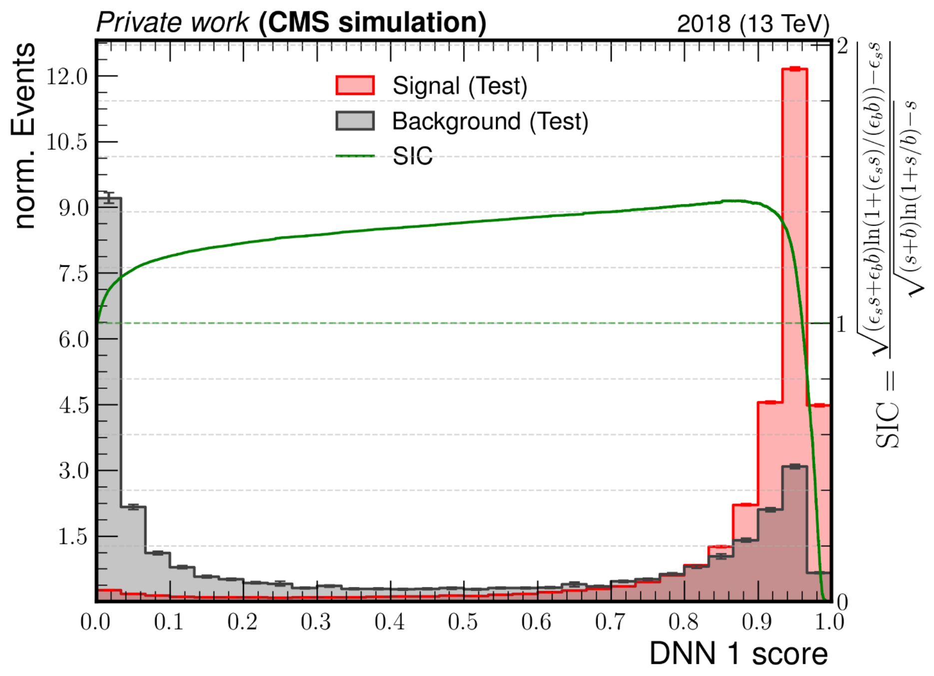
- …atistically independent
- …contamination in CRs
- …ble
- …ncorrelated to M_{T}
- …maximize the significance …ignal region
- …ate X, Y against $\Delta\eta$ to encourage closure in signal and validation plane
- enhance closure in normalization [1] and shape (novelty)

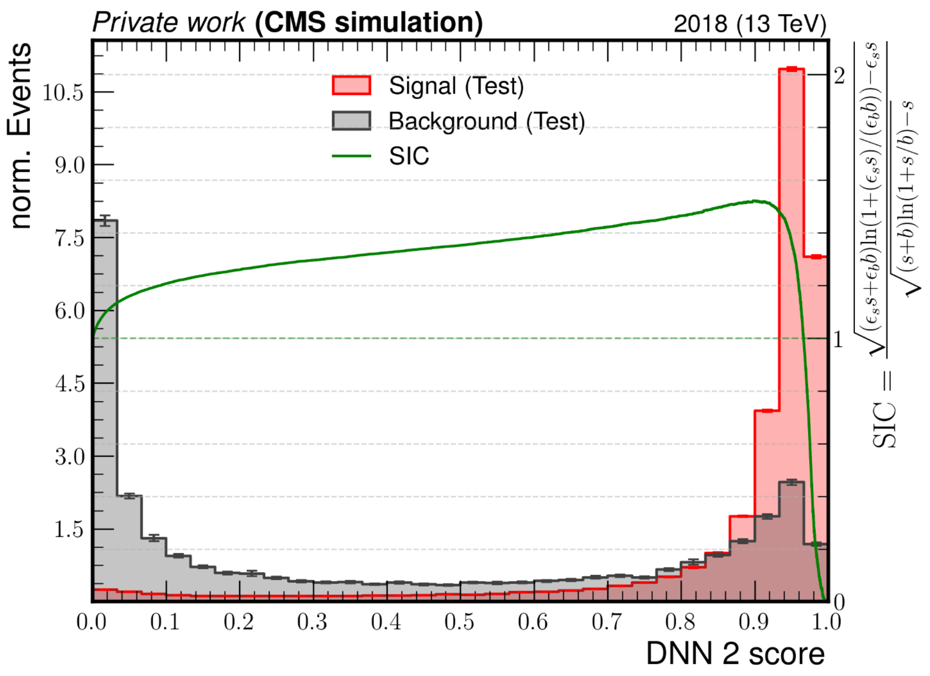
[1] https://arxiv.org/pdf/2506.08826

Results

after inference on the test dataset...

III) Score Distributions in the Signal Plane (SP)



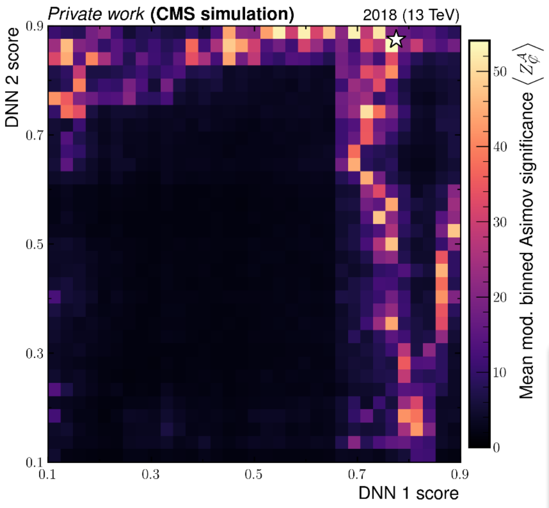


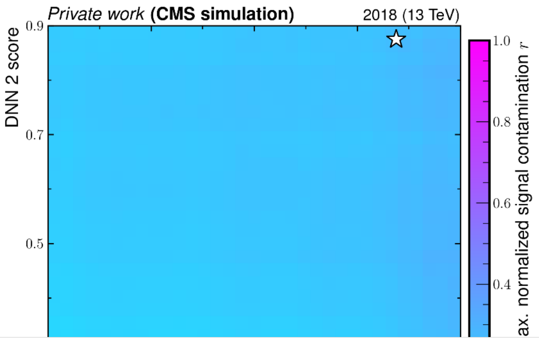
III) Boundary Optimization

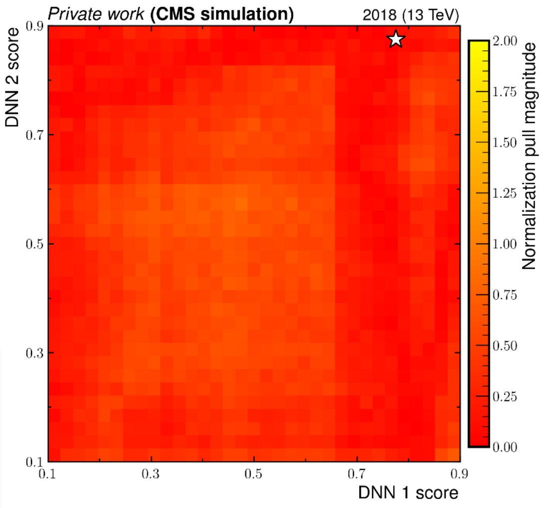
- Goal: Find stable working point for ABCD cuts that **maximizes sensitivity**
- Run **grid search** (33 x 33 in $[0.1, 0.9]^2$), for each point:
  - Build M_T templates in {A, B, C, D}, produce B-only Asimov dataset
  - Perform B-only ABCD fit (CRs only) → Propagate uncertainties to SR prediction
  - Compute metrics:

Metric	**Rule**
Mean mod. binned Asimov significance $\langle Z^{A}_{\not{C}} \rangle$	maximize
SR χ^2 test p-value	$\geq 90\%$ and on a local plateau
Absolute fractional non-closure $\lvert\delta_{\not{C}}\rvert$	$\leq 5\%$
$\lvert$Pull$\rvert$ (normalization)	≤ 2.0
Max. normalized signal contamination $r_{\max}$	≤ 0.35
WP bounds	$X, Y \in [0.1, 0.9]$

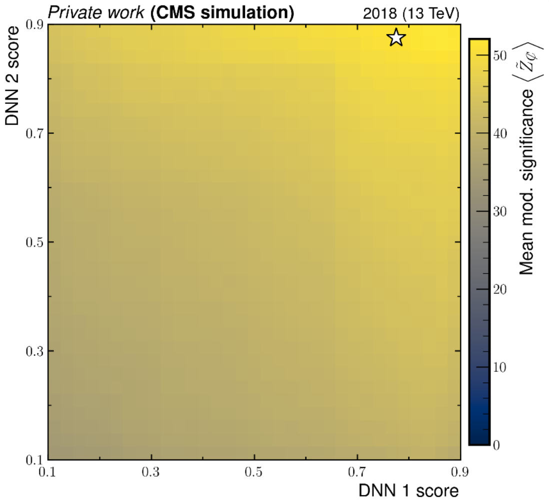
III) Boundary Optimization – Grid Scan

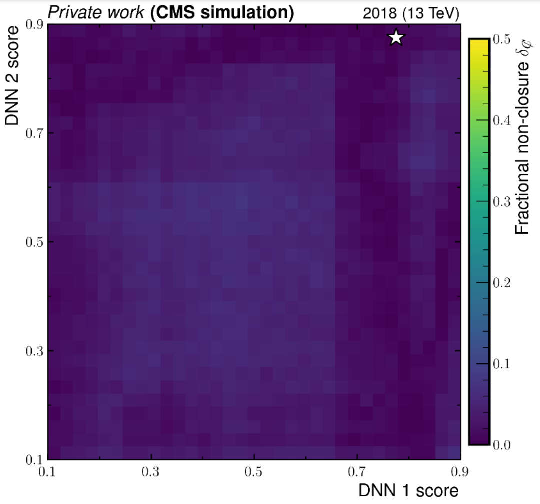


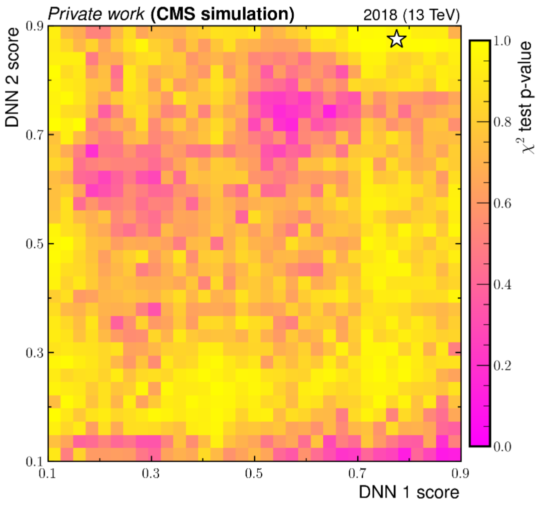




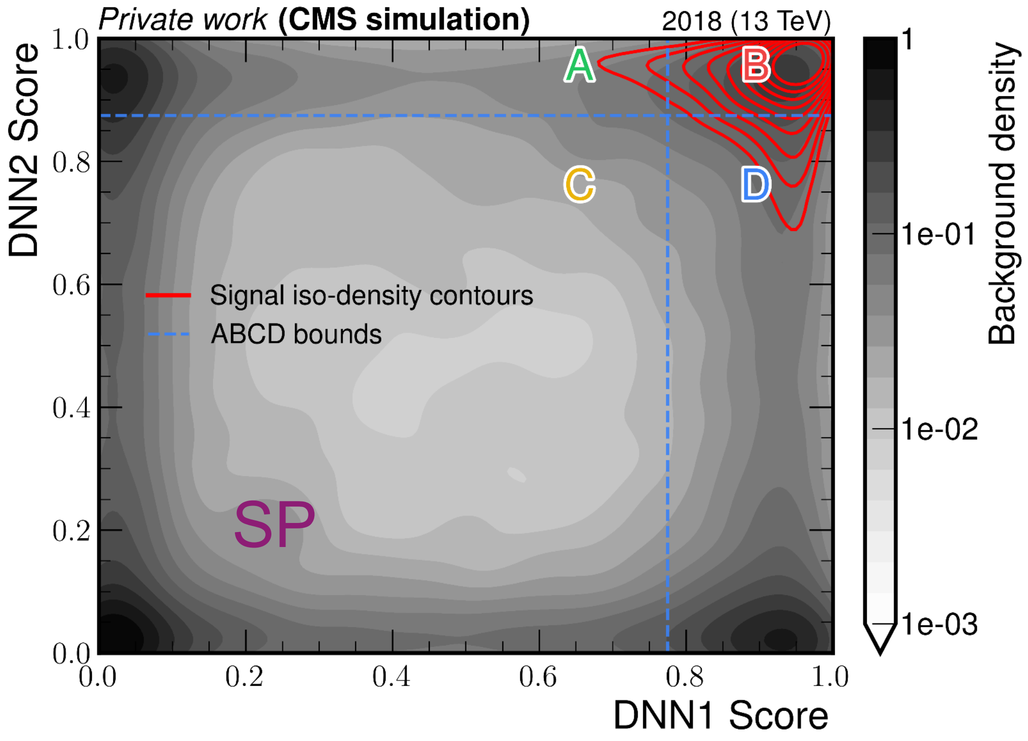
Metric	Rule	Result at WP
Mean mod. binned Asimov significance $\langle Z^A_{\not{C}} \rangle$	maximize	54.1
SR χ^2 test p-value	$\geq 90\%$ and on a local plateau	96.8%
Absolute fractional non-closure $\lvert\delta_{\not{C}}\rvert$	$\leq 5\%$	0.53%
$\lvert$Pull$\rvert$ (normalization)	≤ 2.0	0.0053
Max. normalized signal contamination $r_{\max}$	≤ 0.35	0.259
WP bounds	$X, Y \in [0.1, 0.9]$	$(0.775, 0.875)$

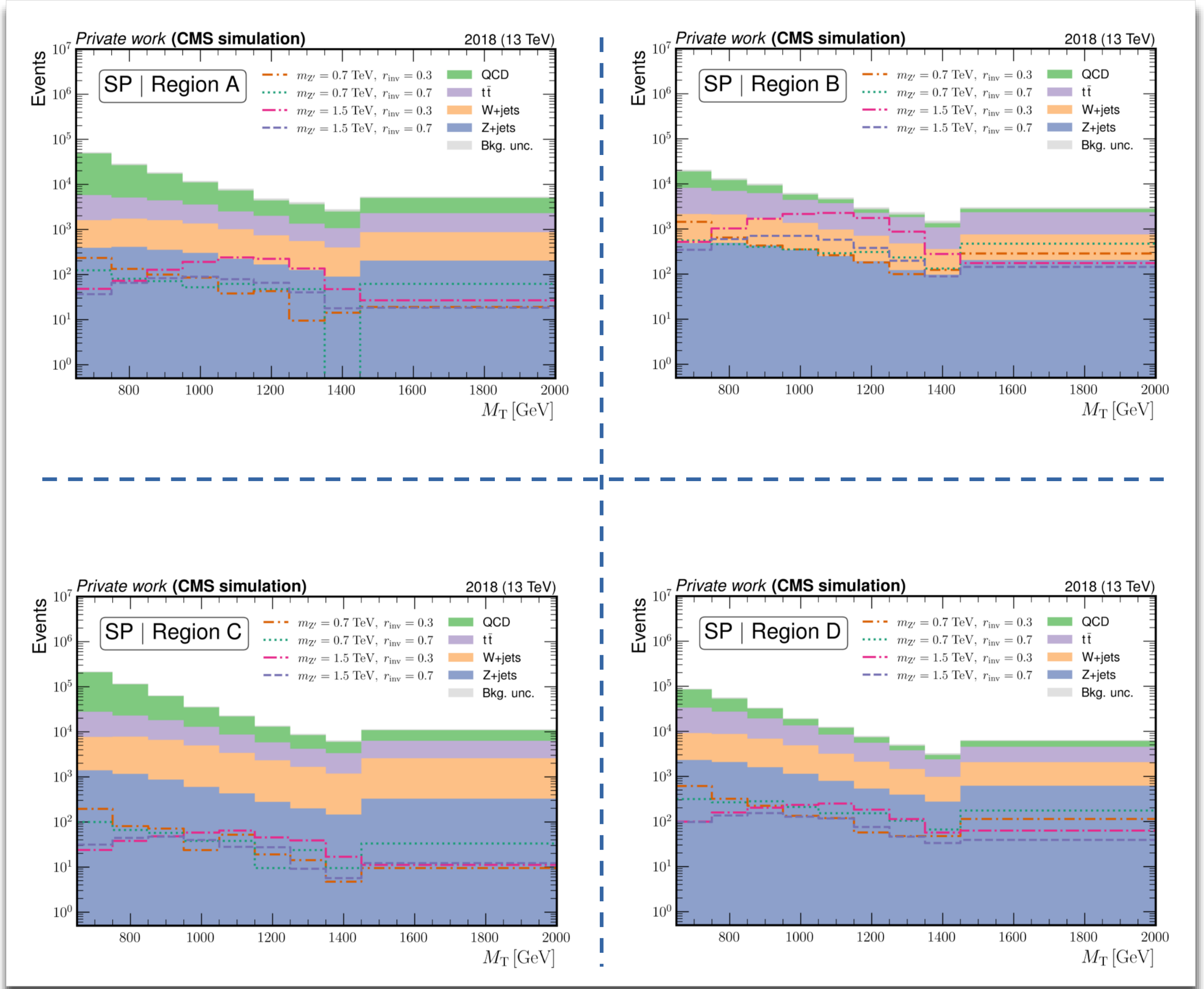






III) Boundary Optimization – Applying the Working Point

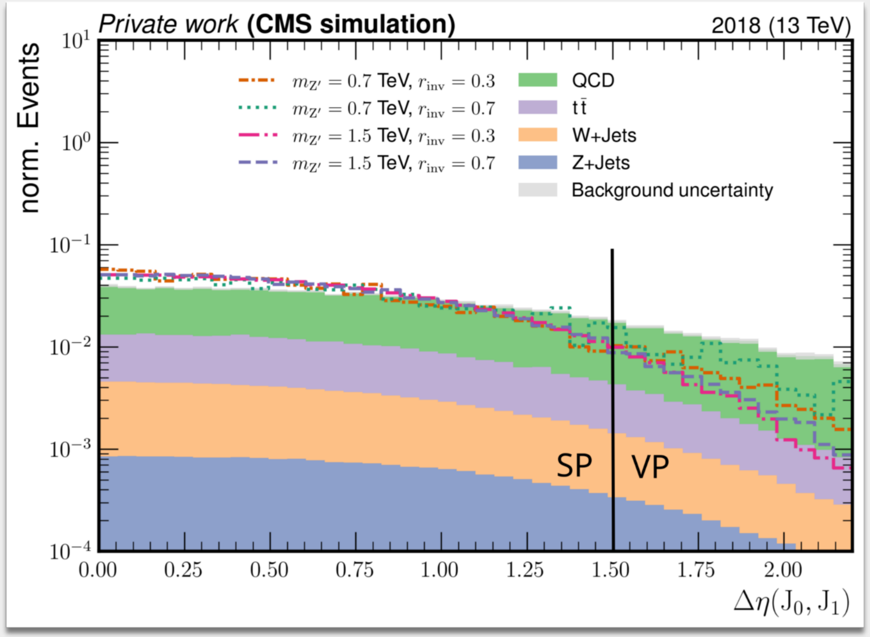




III) Validation

- To demonstrate: stable, unbiased background prediction at chosen WP before unblinding
- Key principles - Validate:
  - ABCD closure
  - Stability against variations of ABCD boundaries
  - Robustness to signal contamination
- Progress from pseudo-data (SP, VP) to data (VP)

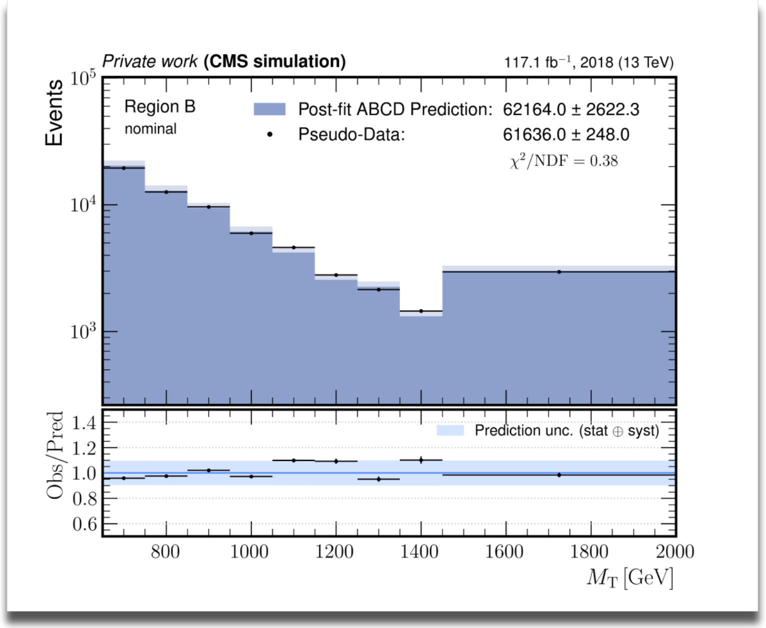
Reminder:



(Distributions after applying preselection)

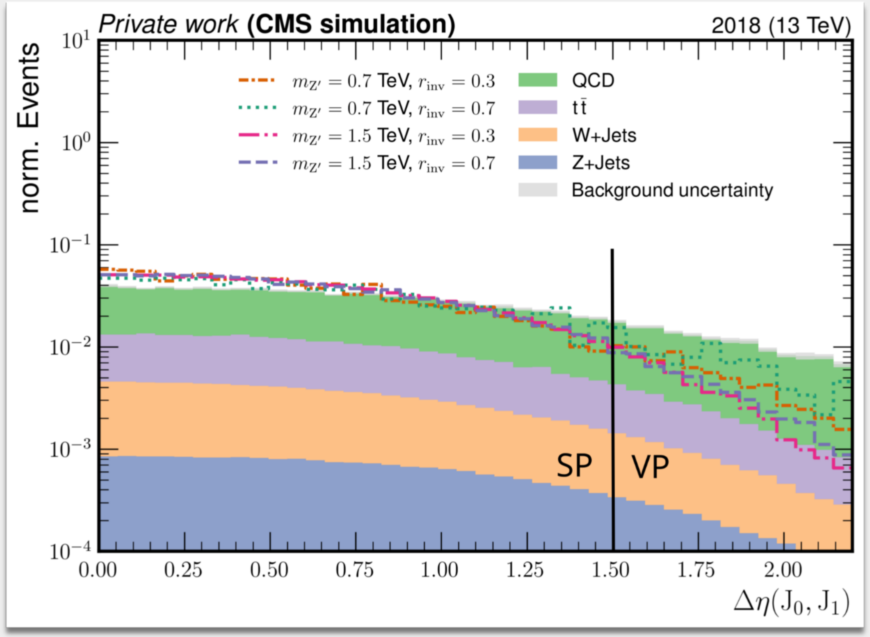
III) Validation

- To demonstrate: stable, unbiased background prediction at chosen WP before unblinding
- Key principles - Validate:
  - ABCD closure
  - Stability against variations of ABCD boundaries
  - Robustness to signal contamination
- Progress from pseudo-data (SP, VP) to data (VP)



→ Derive (per-bin) non-closure uncertainties for stat. inference

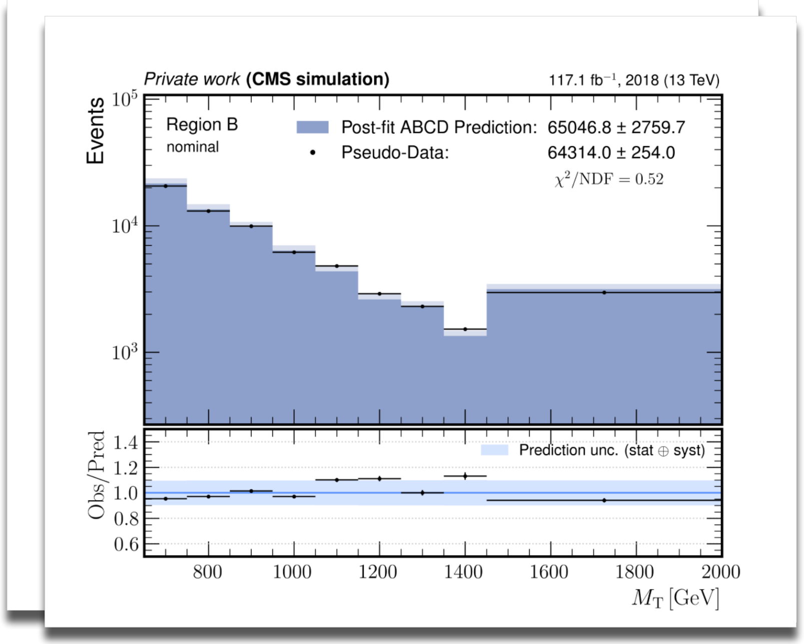
Reminder:



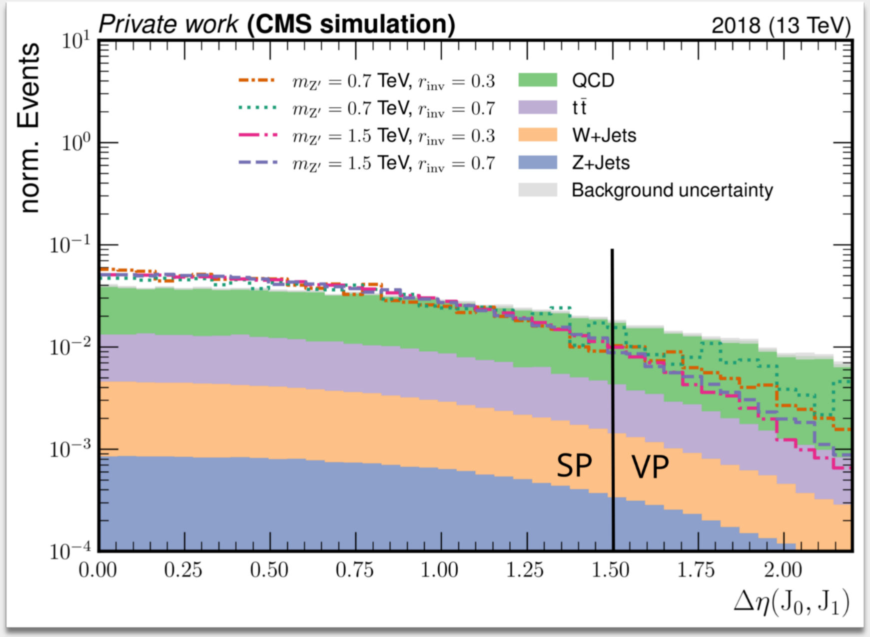
(Distributions after applying preselection)

III) Validation

- <u>To demonstrate:</u> stable, unbiased background prediction at chosen WP before unblinding
- Key principles - Validate:
  - ABCD closure
  - Stability against variations of ABCD boundaries
  - Robustness to signal contamination
- Progress from pseudo-data (SP, VP) to data (VP)



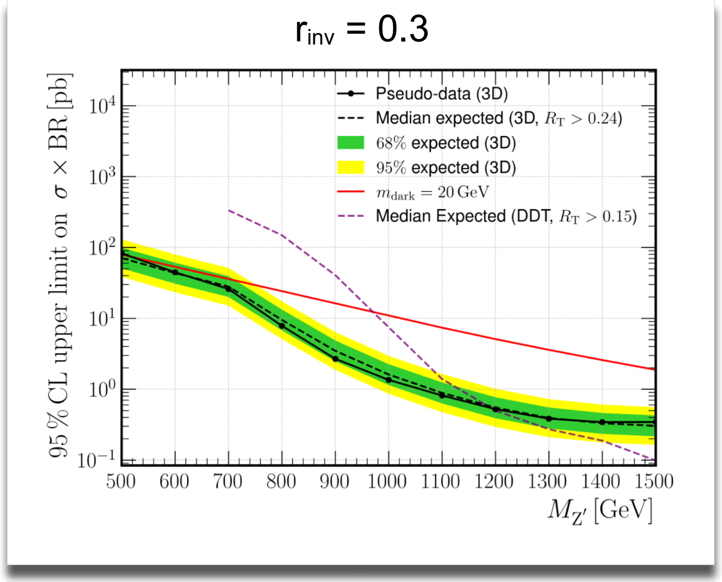
Reminder:

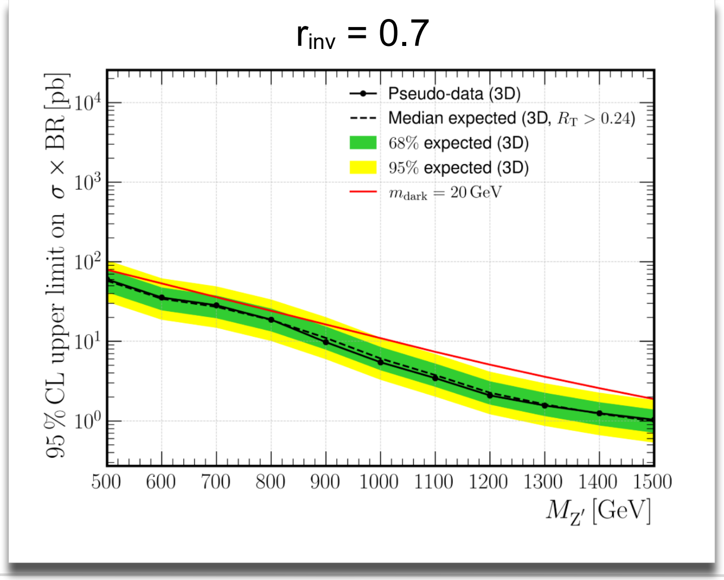


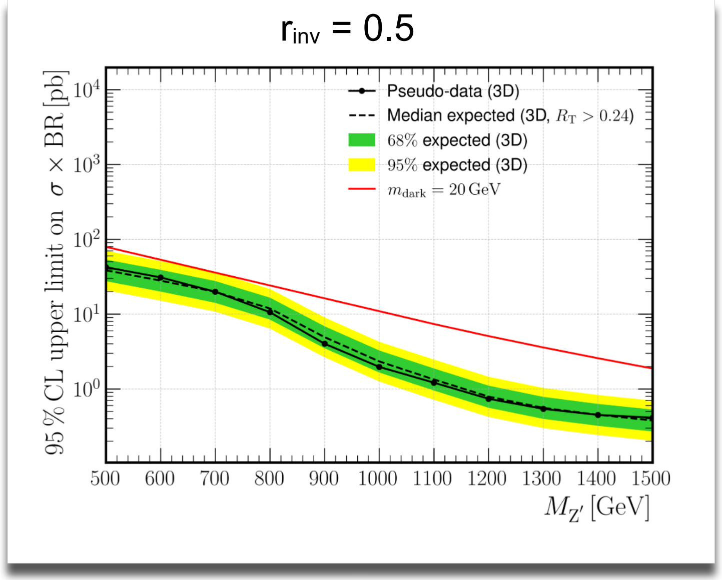
(Distributions after applying preselection)

IV) Statistical Inference: Expected Limits

- Performed with CMS Combine via CLs method
- Using the profile-likelihood test statistic & the asymptotic approximation
- Combined S+B fit across regions {A,B,C,D} per bin
- Include per-bin non-closure systematics for background
- Signal systematics to be added



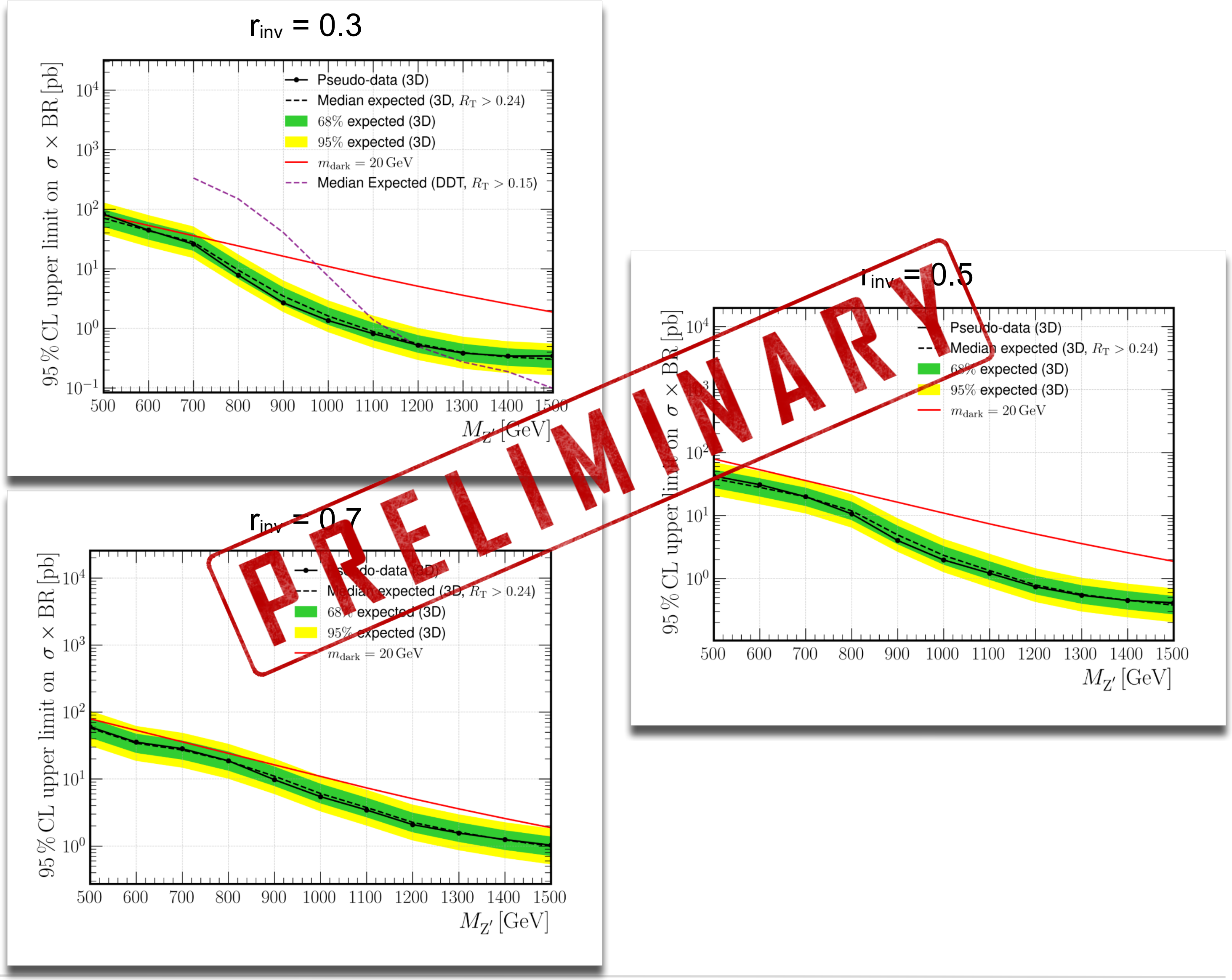




IV) Statistical Inference: Expected Limits

- Performed with CMS Combine via CLs method
- Using the profile-likelihood test statistic & the asymptotic approximation
- Combined S+B fit across regions {A,B,C,D} per bin
- Include per-bin non-closure systematics for background
- Signal systematics to be added

WORK IN PROGRESS



Future Directions

of the ABCD method...

Rethinking ABCD Automation

Karlsruhe Institute of Technology

$\mathcal{L}(\theta;\lambda) = \underbrace{\lambda^{(X)}_{\mathrm{BCE}}\mathcal{L}^{(X)}_{\mathrm{BCE}} + \lambda^{(Y)}_{\mathrm{BCE}}\mathcal{L}^{(Y)}_{\mathrm{BCE}}}_{s\text{–}b\text{ discrimination}}$

$+ \underbrace{\lambda^{(X,Y)}_{\mathrm{DisCo}}\widehat{\mathrm{dCorr}}_b(X,Y)}_{\text{decorrelation of } X \text{ and } Y}$

$+ \lambda^{(X,M_T)}_{\mathrm{DisCo}}\widehat{\mathrm{dCorr}}_b(X,M_T) + \lambda^{(Y,\ldots}_{\mathrm{Dis}\ldots}$ … decorrelation of $\{X,Y\}$ …

$+ \lambda^{(X,\Delta\eta)}_{\mathrm{DisCo}}\widehat{\mathrm{dCorr}}_b(X,\Delta\eta) + \lambda^{(Y,\Delta\ldots}_{\mathrm{DisC}\ldots}$ … decorrelation of $\{X,Y\}$ a…

$+ \underbrace{\sum_{k\in\{\mathrm{SP,VP}\}} \lambda^{(k)}_{\mathcal{C}}\mathcal{C}^b_k}_{\text{normalization closure}}$

$+ \underbrace{\sum_{k\in\{\mathrm{SP,VP}\}}\sum_{i=1}^{N_{\mathrm{bins}}} \lambda^{(k)}_{\mathcal{C},i}\mathcal{C}^b_{k,i}}_{\text{shape-level closure}}$.

Modified Differential Multiplier Method (MDMM)

- keep BCE sum as primary optimization target
- treat other loss terms as inequality constraints $\mathcal{L}_j(\theta) \leq \epsilon_j$

$$\phi_j(\theta, s_j) = \epsilon_j - \mathcal{L}_j(\theta) - s_j^2\,,$$

- Augmented Lagrangian formulation:

$$\tilde{\mathcal{L}}(\theta,\lambda,s) = \mathcal{L}^{(X)}_{\mathrm{BCE}} + \mathcal{L}^{(Y)}_{\mathrm{BCE}} + \sum_{j\in\mathcal{J}} \alpha_j\left(-\tilde{\lambda}_j\,\phi_j(\theta,s_j) + \tfrac{\rho_j}{2}\,\phi_j(\theta,s_j)^2\right),$$

- previously fixed loss weights λ become trainable multipliers
- update network parameters by gradient descent, multipliers by dual gradient ascent

Closure conditions

…atistically independent
…contamination in CRs
…ble
…ncorrelated to M_T

…maximize the significance
…ignal region

…ate X, Y against $\Delta\eta$ to encourage closure in signal and validation plane

- enhance closure in normalization [1] and shape (novelty)

- How to find optimal target values for constrained losses?
- Training objective =/= Analysis objective
- Distance Correlation loss limits batch size during training (running into VRAM limits)

  → Closure & DisCo losses will be biased estimators
- Systematic non-closure uncertainties dominate in SVJ Scouting phase space (too many background events)
- Can we consider systematic uncertainties during training?

Rethinking ABCD Automation

Karlsruhe Institute of Technology

$\mathcal{L}(\theta;\lambda) = \lambda^{(X)}_{\mathrm{BCE}}\,\mathcal{L}^{(X)}_{\mathrm{BCE}} + \lambda^{(Y)}_{\mathrm{BCE}}\,\mathcal{L}^{(Y)}_{\mathrm{BCE}}$ (s−b discrimination)

$+\,\lambda^{(X,Y)}_{\mathrm{DisCo}}\,\widehat{\mathrm{dCorr}}_b(X,Y)$ (decorrelation of X and Y)

$+\,\lambda^{(X,M_T)}_{\mathrm{DisCo}}\,\widehat{\mathrm{dCorr}}_b(X,M_T) + \ldots$ (decorrelation of {X,Y} ...)

$+\,\lambda^{(X,\Delta\eta)}_{\mathrm{DisCo}}\,\widehat{\mathrm{dCorr}}_b(X,\Delta\eta) + \ldots$ (decorrelation of {X,Y} ...)

$+\sum_{k\in\{\mathrm{SP,VP}\}} \lambda^{(k)}_{\mathcal{C}}\,\mathcal{C}^b_k$ (normalization closure)

$+\sum_{k\in\{\mathrm{SP,VP}\}}\sum_{i=1}^{N_{\mathrm{bins}}} \lambda^{(k)}_{\mathcal{C},i}\,\mathcal{C}^b_{k,i}$ (shape-level closure)

Modified Differential Multiplier Method (MDMM)

- keep BCE sum as primary optimization target
- treat other loss terms as inequality constraints $\mathcal{L}_j(\theta) \le \epsilon_j$

$$\phi_j(\theta, s_j) = \epsilon_j - \mathcal{L}_j(\theta) - s_j^2\,,$$

- Augmented Lagrangian formulation:

$$\tilde{\mathcal{L}}(\theta,\lambda,s) = \mathcal{L}^{(X)}_{\mathrm{BCE}} + \mathcal{L}^{(Y)}_{\mathrm{BCE}} + \sum_{j\in\mathcal{J}} \alpha_j \left(-\tilde{\lambda}_j\,\phi_j(\theta,s_j) + \tfrac{\rho_j}{2}\,\phi_j(\theta,s_j)^2\right),$$

- previously fixed loss weights λ become trainable multipliers
- update network parameters by gradient descent, multipliers by dual gradient ascent

Closure conditions

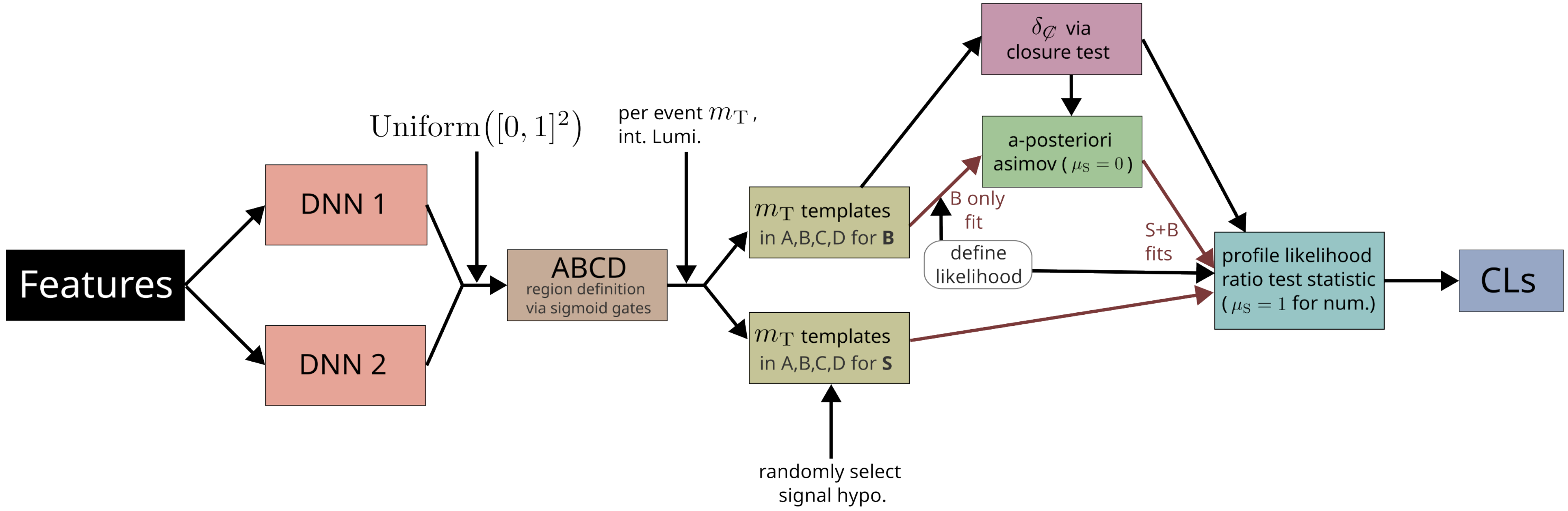
...atistically independent
...contamination in CRs
...ble
...ncorrelated to M_T
...maximize the significance ...ignal region
...ate X, Y against $\Delta\eta$ to encourage closure in signal and validation plane
- enhance closure in normalization [1] and shape (novelty)

- How to find optimal target values for constrained losses?
- Training objective =/= Analysis objective
- Distance Correlation loss limits batch size during training (running into VRAM limits)

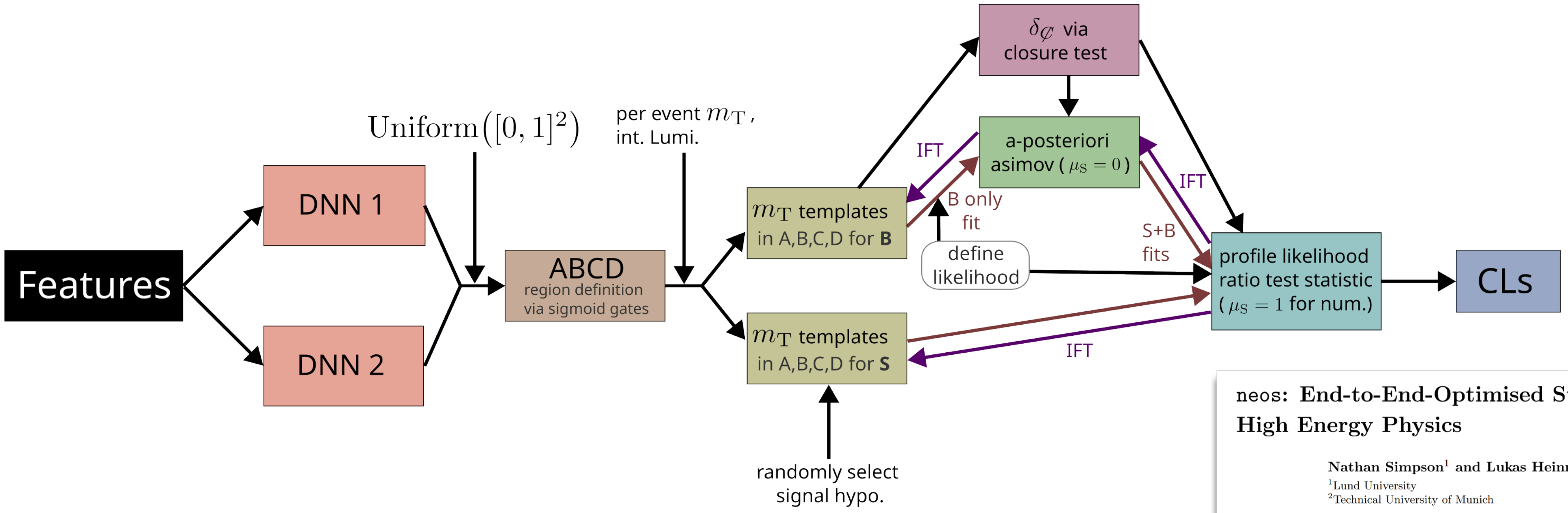
  → Closure & DisCo losses will be biased estimators
- Systematic non-closure uncertainties dominate in SVJ Scouting phase space (too many background events)
- Can we consider systematic uncertainties during training?

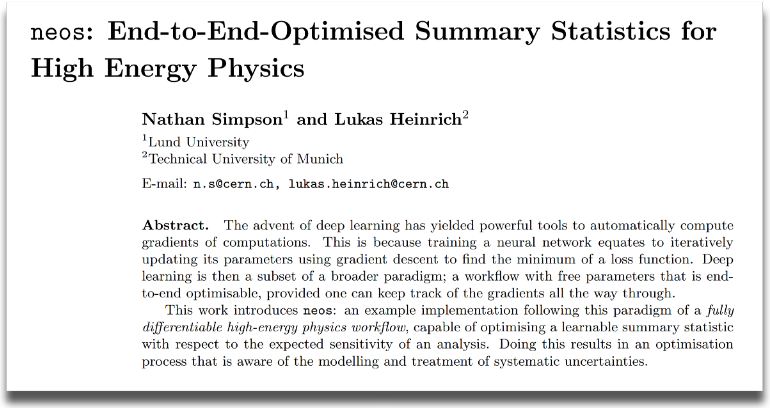
➡ Replace BCE with with a loss derived from our analysis objective! (CLs)

Toward an End-to-End optimized binned ABCD...



Toward an End-to-End optimized binned ABCD...





neos: End-to-End-Optimised Summary Statistics for High Energy Physics

Nathan Simpson[1] and Lukas Heinrich[2]
[1]Lund University
[2]Technical University of Munich
E-mail: n.s@cern.ch, lukas.heinrich@cern.ch

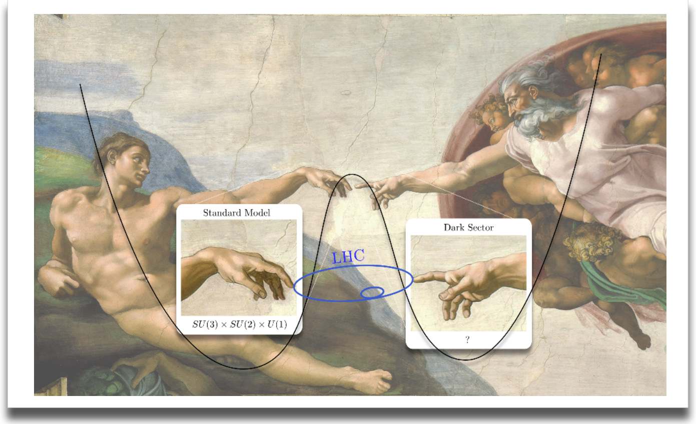
Abstract. The advent of deep learning has yielded powerful tools to automatically compute gradients of computations. This is because training a neural network equates to iteratively updating its parameters using gradient descent to find the minimum of a loss function. Deep learning is then a subset of a broader paradigm; a workflow with free parameters that is end-to-end optimisable, provided one can keep track of the gradients all the way through.

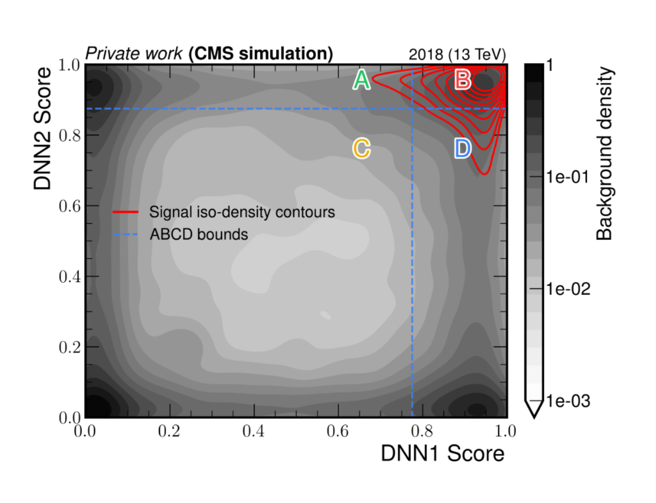
This work introduces neos: an example implementation following this paradigm of a *fully differentiable high-energy physics workflow*, capable of optimising a learnable summary statistic with respect to the expected sensitivity of an analysis. Doing this results in an optimisation process that is aware of the modelling and treatment of systematic uncertainties.

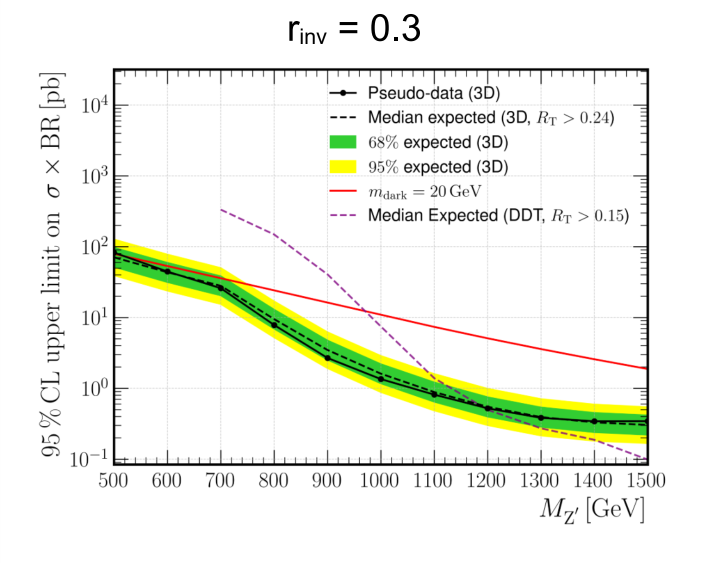
arXiv:2203.05570

Summary

- **Analysis pipeline established** for first CMS Scouting search for semi-visible jets
- **Supervised ML assisted strategy:** LundNet jet tagging & 3D-ABCDisCoTEC event tagging + ABCD
- **Novel shape-level generalization** of the ABCDisCoTEC approach (suitable for Scouting)
- **Validation strategy well-advanced**, closure-testing on data (validation plane) soon!
- **Improved sensitivity** at low mediator masses
- Foundations laid for **end-to-end, analysis-aware ABCD methods**

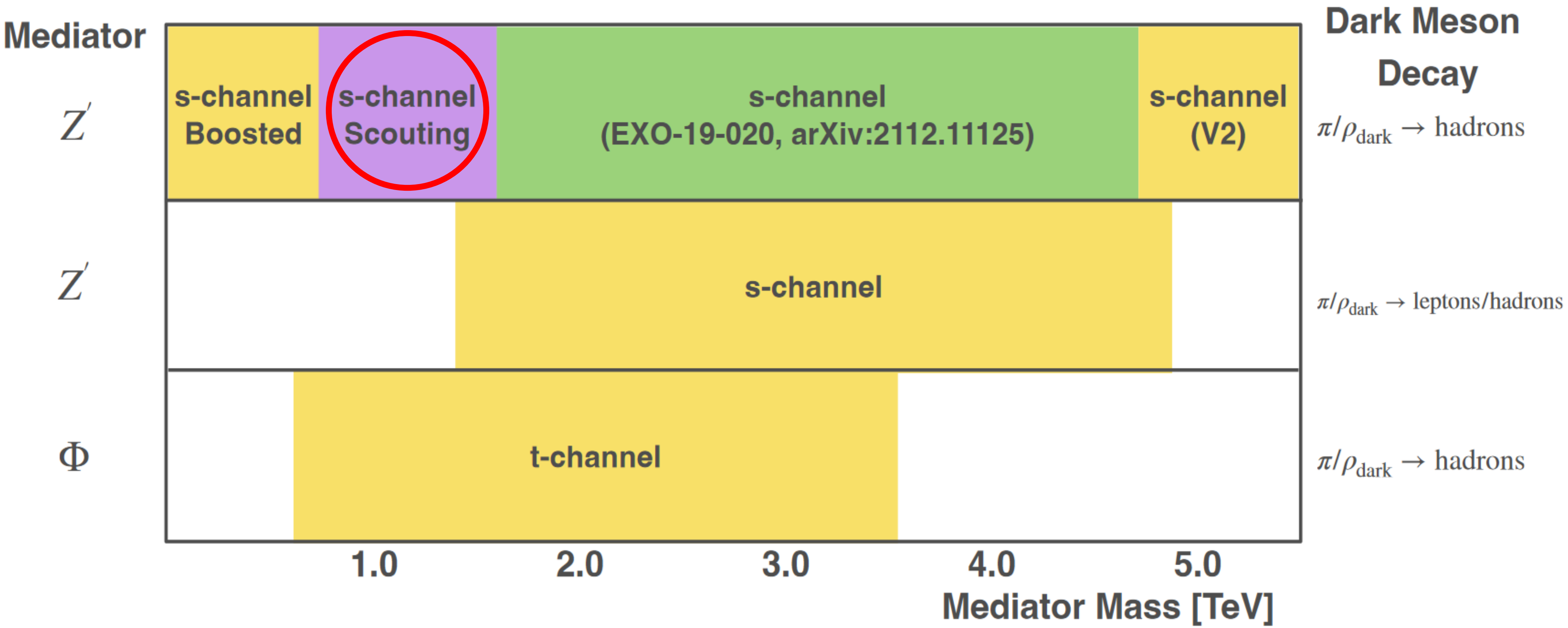






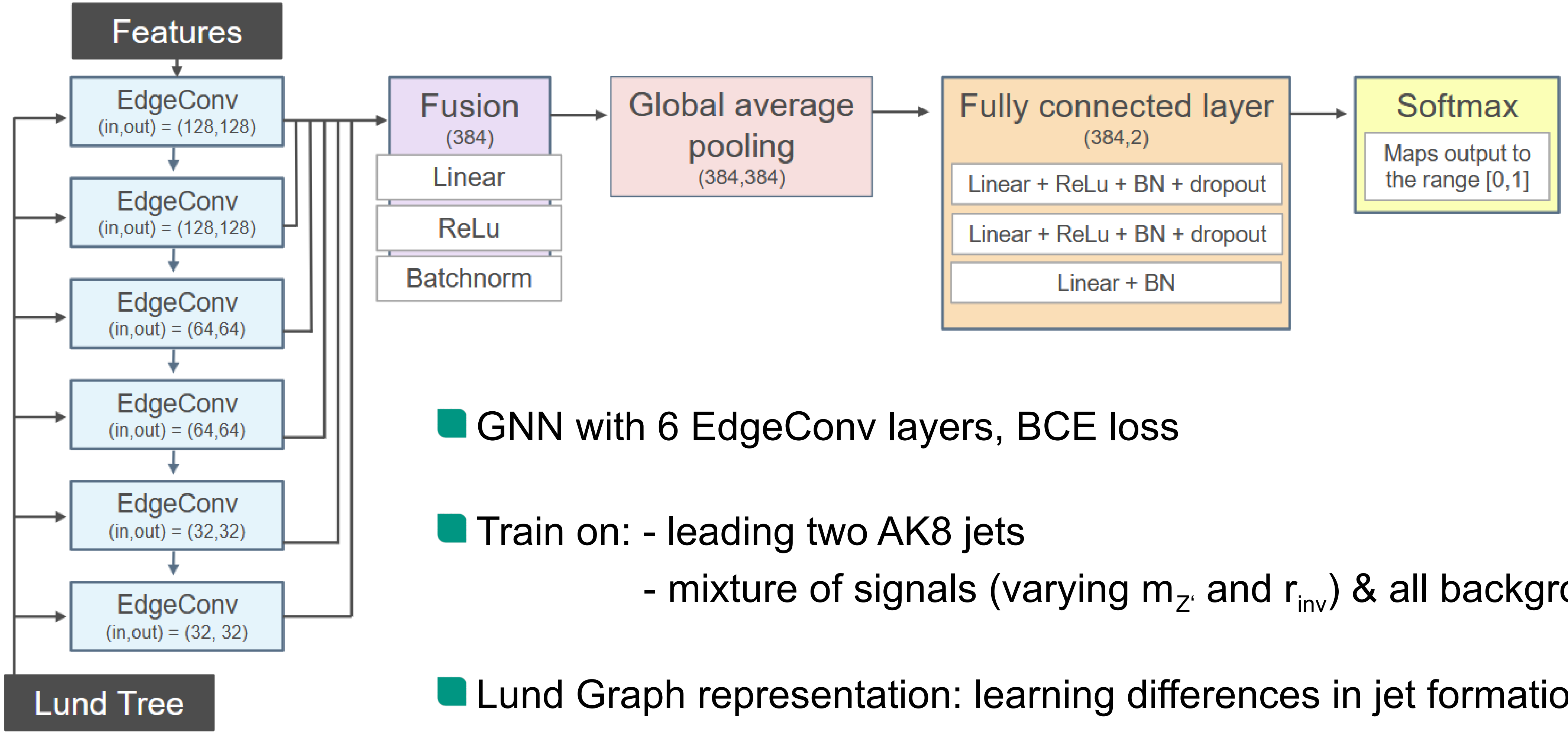
Backup

SVJ Search Efforts at CMS – An Overview



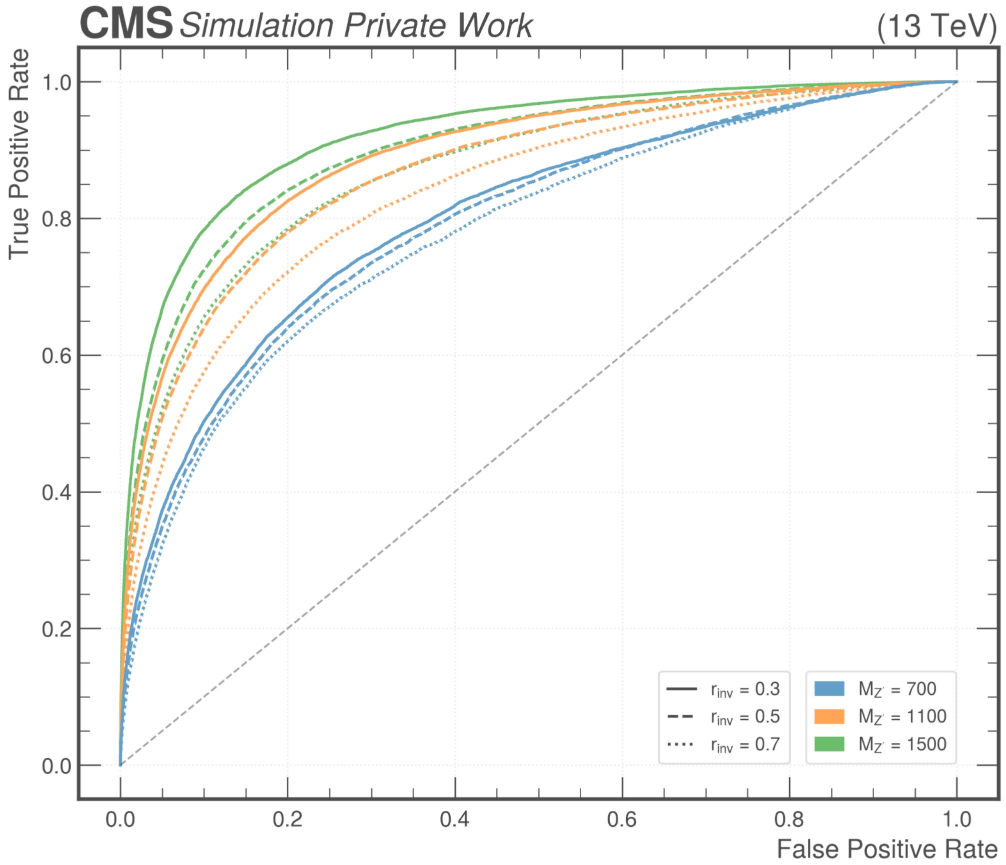
Scanned model parameters: $\boldsymbol{m_{Z'}}$, r_{inv}, (m_{dark}) ...

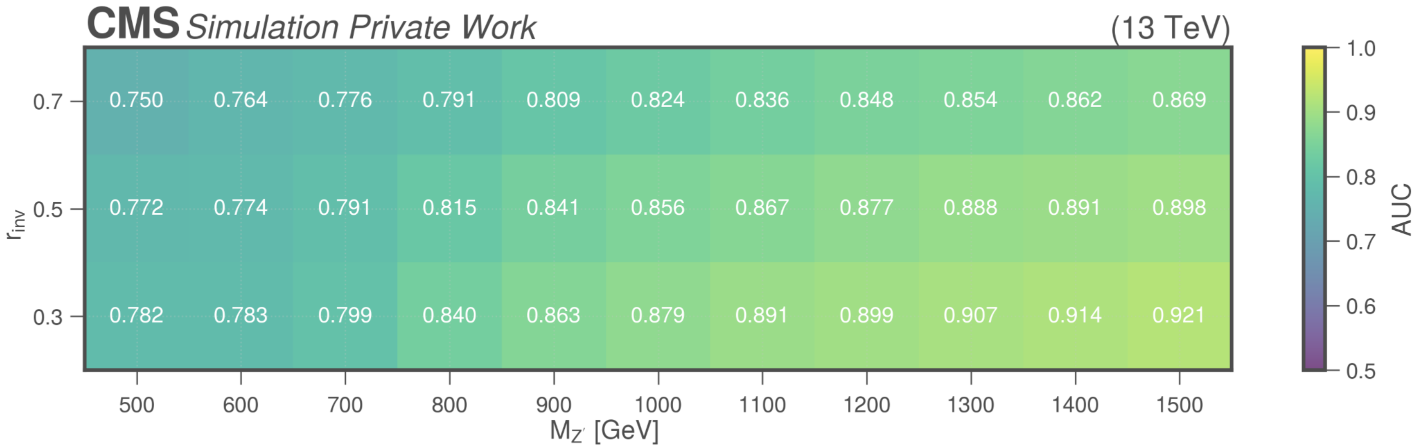
II) Supervised SVJ Tagging via LundNet



- GNN with 6 EdgeConv layers, BCE loss
- Train on: - leading two AK8 jets
  - mixture of signals (varying $m_{Z'}$ and r_{inv}) & all backgrounds
- Lund Graph representation: learning differences in jet formation

II) Supervised SVJ Tagging via LundNet



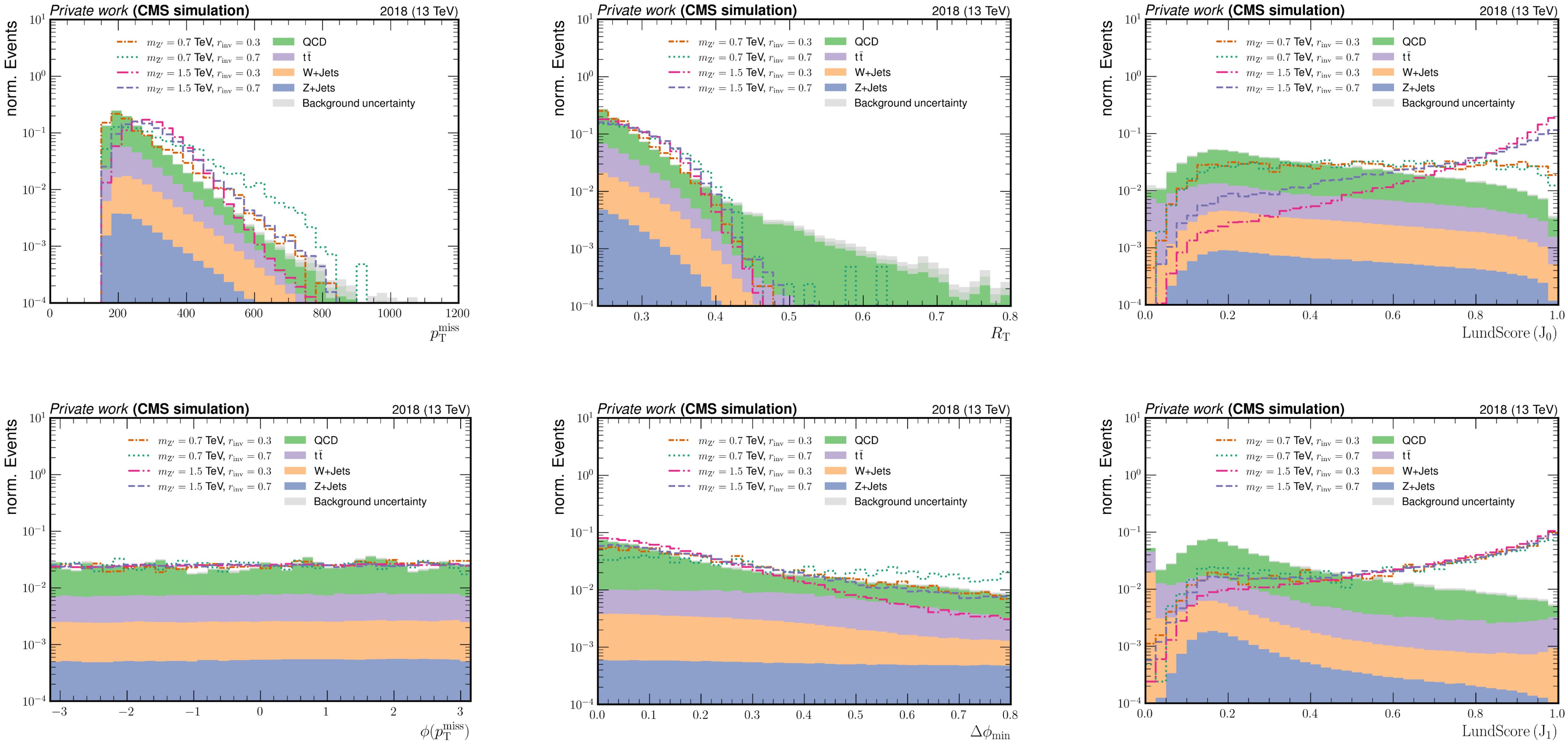


r_{inv} \ $M_{Z'}$ [GeV]	500	600	700	800	900	1000	1100	1200	1300	1400	1500
0.7	0.750	0.764	0.776	0.791	0.809	0.824	0.836	0.848	0.854	0.862	0.869
0.5	0.772	0.774	0.791	0.815	0.841	0.856	0.867	0.877	0.888	0.891	0.898
0.3	0.782	0.783	0.799	0.840	0.863	0.879	0.891	0.899	0.907	0.914	0.921

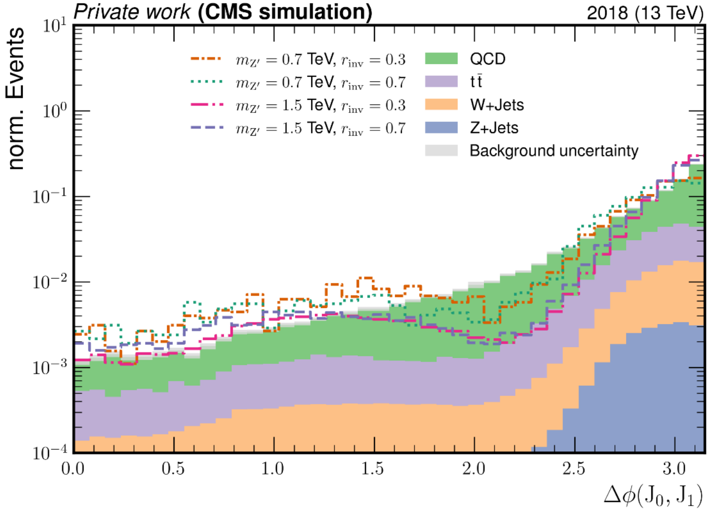
Overall good performance, with AUCs degrading for low $m_{Z'}$ and high r_{inv}

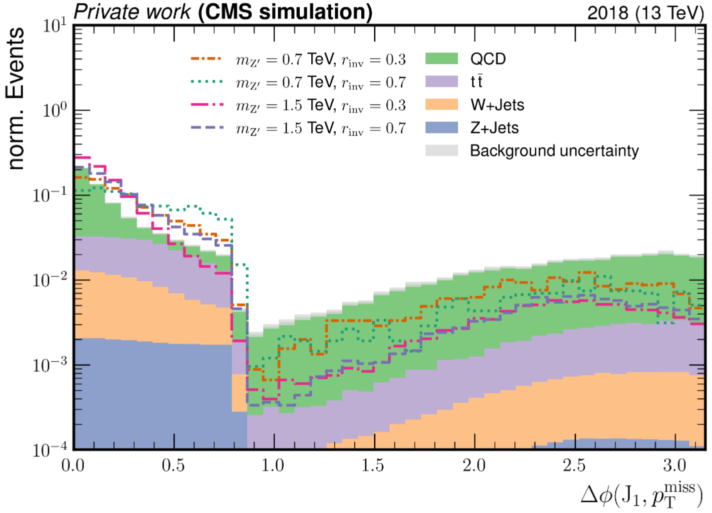
→ This is due to signal jets becoming more QCD-like at low $m_{Z'}$ / very invisible at high r_{inv}

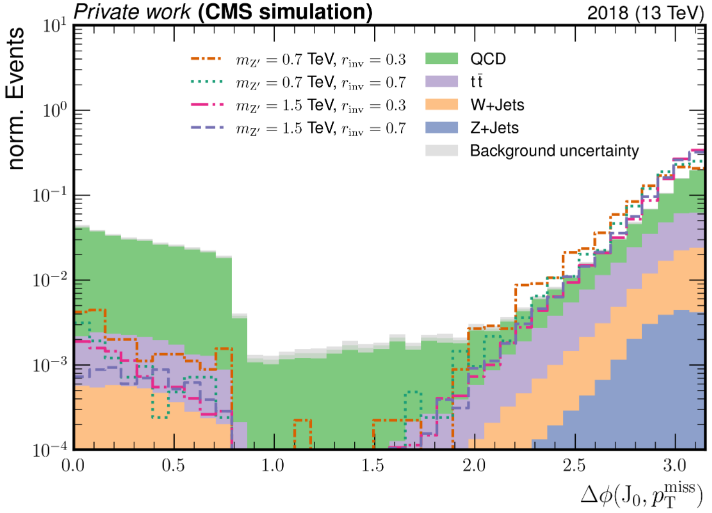
III) 3D-ABCDisCoTEC Method – Input Features (i)

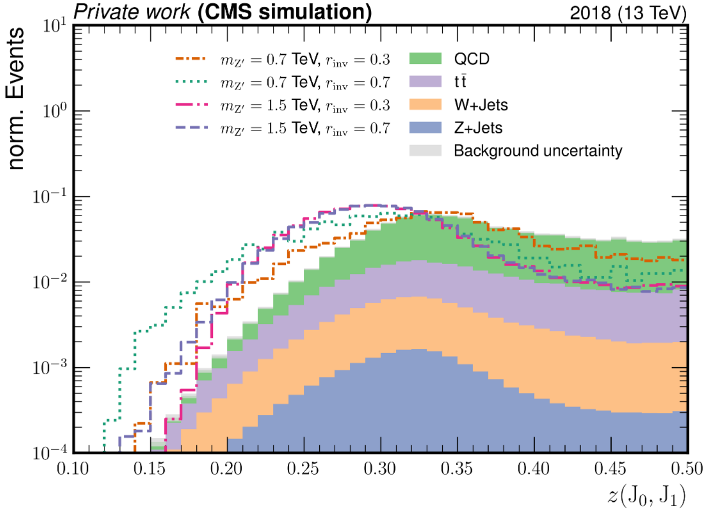


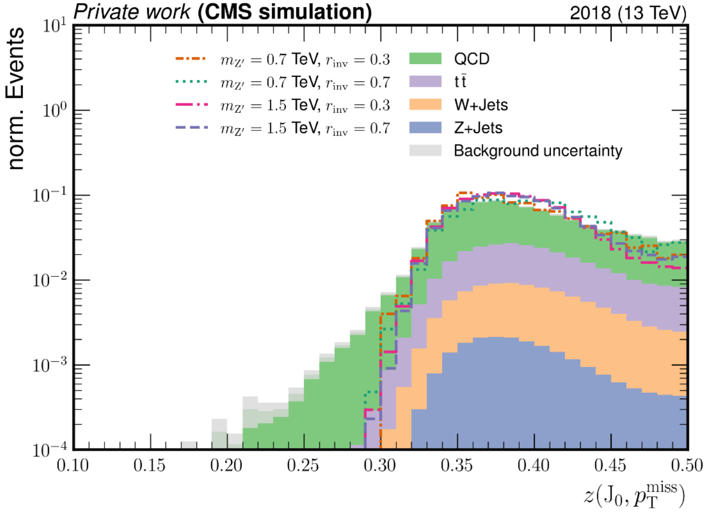
III) 3D-ABCDisCoTEC Method – Input Features (ii)

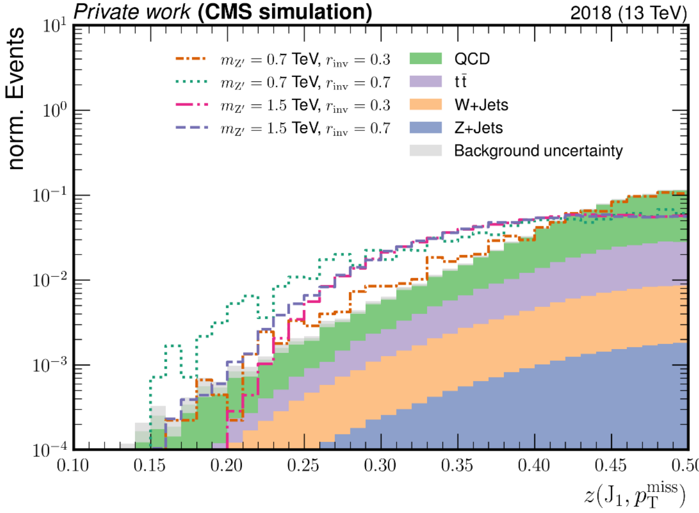






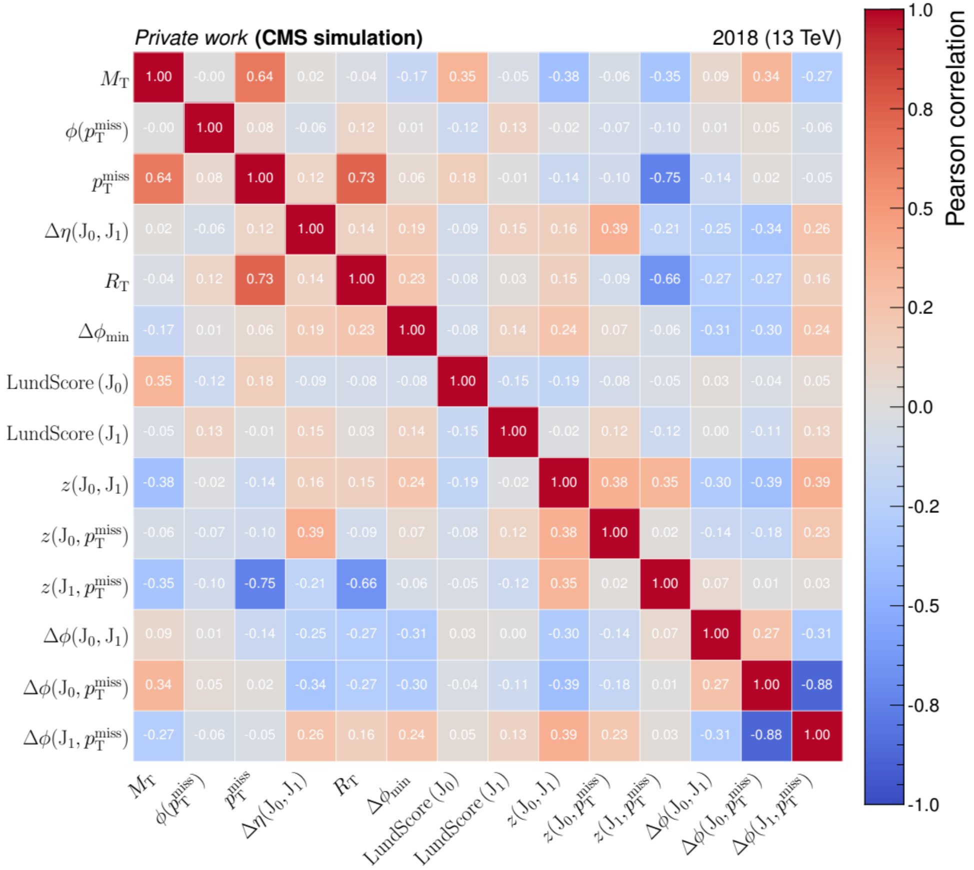


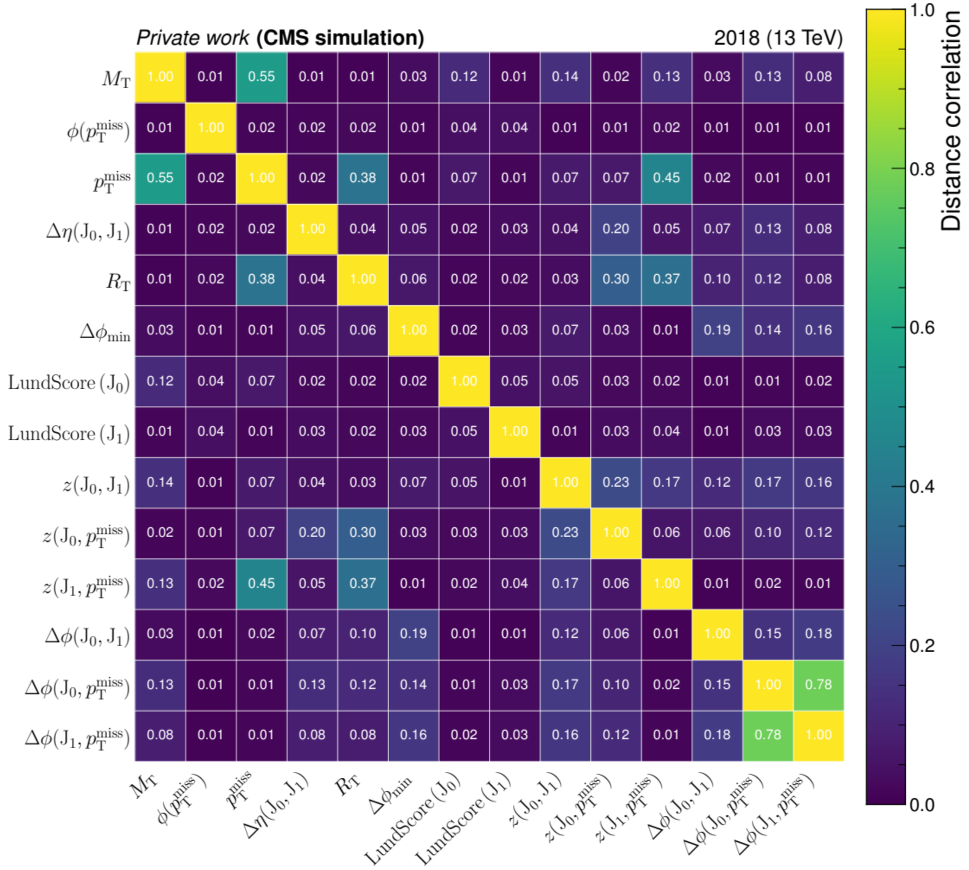




III) 3D-ABCDisCoTEC Method – Corr. (Features)

Karlsruhe Institute of Technology





III) 3D-ABCDisCoTEC: Loss details (i)

$$\mathcal{L}(\theta;\lambda) = \underbrace{\lambda^{(X)}_{\mathrm{BCE}}\,\mathcal{L}^{(X)}_{\mathrm{BCE}} + \lambda^{(Y)}_{\mathrm{BCE}}\,\mathcal{L}^{(Y)}_{\mathrm{BCE}}}_{s-b\text{ discrimination}}$$

$$+ \underbrace{\lambda^{(X,Y)}_{\mathrm{DisCo}}\,\widehat{\mathrm{dCorr}}_b(X,Y)}_{\text{decorrelation of } X \text{ and } Y}$$

$$+ \underbrace{\lambda^{(X,M_\mathrm{T})}_{\mathrm{DisCo}}\,\widehat{\mathrm{dCorr}}_b(X,M_\mathrm{T}) + \lambda^{(Y,M_\mathrm{T})}_{\mathrm{DisCo}}\,\widehat{\mathrm{dCorr}}_b(Y,M_\mathrm{T})}_{\text{decorrelation of } \{X,Y\} \text{ and } M_\mathrm{T}}$$

$$+ \underbrace{\lambda^{(X,\Delta\eta)}_{\mathrm{DisCo}}\,\widehat{\mathrm{dCorr}}_b(X,\Delta\eta) + \lambda^{(Y,\Delta\eta)}_{\mathrm{DisCo}}\,\widehat{\mathrm{dCorr}}_b(Y,\Delta\eta)}_{\text{decorrelation of } \{X,Y\} \text{ and } \Delta\eta}$$

$$+ \underbrace{\sum_{k\in\{\mathrm{SP},\mathrm{VP}\}} \lambda^{(k)}_{\not{C}}\,\not{C}^b_k}_{\text{normalization closure}}$$

$$+ \underbrace{\sum_{k\in\{\mathrm{SP},\mathrm{VP}\}} \sum_{i=1}^{N_\mathrm{bins}} \lambda^{(k)}_{\not{C},i}\,\not{C}^b_{k,i}}_{\text{shape-level closure}} .$$

$$\mathcal{L}^{(X)}_{\mathrm{BCE}} = -\frac{1}{\sum_i w_i}\sum_{i=1}^{N} w_i\Big[y_i \ln\sigma(X_i) + (1-y_i)\ln(1-\sigma(X_i))\Big],$$

$$\mathcal{L}^{(Y)}_{\mathrm{BCE}} = -\frac{1}{\sum_i w_i}\sum_{i=1}^{N} w_i\Big[y_i \ln\sigma(Y_i) + (1-y_i)\ln(1-\sigma(Y_i))\Big],$$

$$\sigma(u) = \frac{1}{1+e^{-u}}$$

III) 3D-ABCDisCoTEC: Loss details (ii)

Karlsruhe Institute of Technology

$$
\begin{aligned}
\mathcal{L}(\theta;\lambda) = {} & \underbrace{\lambda^{(X)}_{\mathrm{BCE}}\,\mathcal{L}^{(X)}_{\mathrm{BCE}} + \lambda^{(Y)}_{\mathrm{BCE}}\,\mathcal{L}^{(Y)}_{\mathrm{BCE}}}_{s-b\ \text{discrimination}} \\
& + \underbrace{\lambda^{(X,Y)}_{\mathrm{DisCo}}\,\widehat{\mathrm{dCorr}}_b(X,Y)}_{\text{decorrelation of } X \text{ and } Y} \\
& + \underbrace{\lambda^{(X,M_\mathrm{T})}_{\mathrm{DisCo}}\,\widehat{\mathrm{dCorr}}_b(X,M_\mathrm{T}) + \lambda^{(Y,M_\mathrm{T})}_{\mathrm{DisCo}}\,\widehat{\mathrm{dCorr}}_b(Y,M_\mathrm{T})}_{\text{decorrelation of } \{X,Y\} \text{ and } M_\mathrm{T}} \\
& + \underbrace{\lambda^{(X,\Delta\eta)}_{\mathrm{DisCo}}\,\widehat{\mathrm{dCorr}}_b(X,\Delta\eta) + \lambda^{(Y,\Delta\eta)}_{\mathrm{DisCo}}\,\widehat{\mathrm{dCorr}}_b(Y,\Delta\eta)}_{\text{decorrelation of } \{X,Y\} \text{ and } \Delta\eta} \\
& + \underbrace{\sum_{k\in\{\mathrm{SP},\mathrm{VP}\}} \lambda^{(k)}_{\not{C}}\,\not{C}^{b}_{k}}_{\text{normalization closure}} \\
& + \underbrace{\sum_{k\in\{\mathrm{SP},\mathrm{VP}\}} \sum_{i=1}^{N_\mathrm{bins}} \lambda^{(k)}_{\not{C},i}\,\not{C}^{b}_{k,i}}_{\text{shape-level closure}} .
\end{aligned}
$$

Sample distance correlation losses are implemented by weighted, upward biased V-statistic estimators evaluated on background events in each minibatch [54]. Let $\mathcal{B} = \{i : y_i = 0\}$ denote the background subset in a minibatch, with the number of background events in the mini-batch $n_b = |\mathcal{B}|$, event weights $\{w_i\}_{i\in\mathcal{B}}$ and total background weight in the batch $W_b = \sum_{i\in\mathcal{B}} w_i$. For variables $x, y \in \mathbb{R}$ and $i, j \in \mathcal{B}$, define the distances

$$d^X_{ij} = |x_i - x_j|\,, \tag{5.18}$$

$$d^Y_{ij} = |y_i - y_j|\,. \tag{5.19}$$

The weighted row mean $\bar{d}^{X,(w)}_{i\cdot}$, weighted column mean $\bar{d}^{X,(w)}_{\cdot j}$, and grand mean $\bar{d}^{X,(w)}_{\cdot\cdot}$ are then

$$\bar{d}^{X,(w)}_{i\cdot} = \frac{1}{W_b}\sum_{j\in\mathcal{B}} w_j d^X_{ij}\,, \tag{5.20}$$

$$\bar{d}^{X,(w)}_{\cdot j} = \frac{1}{W_b}\sum_{i\in\mathcal{B}} w_i d^X_{ij}\,, \tag{5.21}$$

$$\bar{d}^{X,(w)}_{\cdot\cdot} = \frac{1}{W_b^2}\sum_{i,j\in\mathcal{B}} w_i w_j d^X_{ij}\,, \tag{5.22}$$

and analogously for d^Y, defining the doubly centered distance matrices

$$D^X_{ij} = d^X_{ij} - \bar{d}^{X,(w)}_{i\cdot} - \bar{d}^{X,(w)}_{\cdot j} + \bar{d}^{X,(w)}_{\cdot\cdot}\,, \tag{5.23}$$

$$D^Y_{ij} = d^Y_{ij} - \bar{d}^{Y,(w)}_{i\cdot} - \bar{d}^{Y,(w)}_{\cdot j} + \bar{d}^{Y,(w)}_{\cdot\cdot}\,. \tag{5.24}$$

The weighted V-statistic estimates of the distance covariance $\widehat{\mathrm{dCov}}^2_b(X,Y)$ and distance variances $\widehat{\mathrm{dVar}}^2_b(\cdot)$ are

$$\widehat{\mathrm{dCov}}^2_b(X,Y) = \frac{1}{W_b^2}\sum_{i,j\in\mathcal{B}} w_i w_j\, D^X_{ij} D^Y_{ij}\,, \tag{5.25}$$

$$\widehat{\mathrm{dVar}}^2_b(X) = \frac{1}{W_b^2}\sum_{i,j\in\mathcal{B}} w_i w_j\, \left(D^X_{ij}\right)^2\,, \tag{5.26}$$

$$\widehat{\mathrm{dVar}}^2_b(Y) = \frac{1}{W_b^2}\sum_{i,j\in\mathcal{B}} w_i w_j\, \left(D^Y_{ij}\right)^2\,, \tag{5.27}$$

and the value entering the loss is the distance correlation $\widehat{\mathrm{dCorr}}^{\,b}(X,Y) \in [0,1]$,

$$\widehat{\mathrm{dCorr}}_b(X,Y) = \frac{\widehat{\mathrm{dCov}}_b(X,Y)}{\sqrt{\widehat{\mathrm{dVar}}_b(X)\,\widehat{\mathrm{dVar}}_b(Y)}}\,, \tag{5.28}$$

III) 3D-ABCDisCoTEC: Loss details (iii)

Karlsruhe Institute of Technology

$$\mathcal{L}(\theta;\lambda) = \underbrace{\lambda_{\text{BCE}}^{(X)}\,\mathcal{L}_{\text{BCE}}^{(X)} + \lambda_{\text{BCE}}^{(Y)}\,\mathcal{L}_{\text{BCE}}^{(Y)}}_{s-b\text{ discrimination}}$$
$$+ \underbrace{\lambda_{\text{DisCo}}^{(X,Y)}\,\widehat{\text{dCorr}}_b(X,Y)}_{\text{decorrelation of } X \text{ and } Y}$$
$$+ \underbrace{\lambda_{\text{DisCo}}^{(X,M_{\text{T}})}\,\widehat{\text{dCorr}}_b(X,M_{\text{T}}) + \lambda_{\text{DisCo}}^{(Y,M_{\text{T}})}\,\widehat{\text{dCorr}}_b(Y,M_{\text{T}})}_{\text{decorrelation of } \{X,Y\} \text{ and } M_{\text{T}}}$$
$$+ \underbrace{\lambda_{\text{DisCo}}^{(X,\Delta\eta)}\,\widehat{\text{dCorr}}_b(X,\Delta\eta) + \lambda_{\text{DisCo}}^{(Y,\Delta\eta)}\,\widehat{\text{dCorr}}_b(Y,\Delta\eta)}_{\text{decorrelation of } \{X,Y\} \text{ and } \Delta\eta}$$
$$+ \underbrace{\sum_{k\in\{\text{SP},\text{VP}\}} \lambda_{\mathcal{C}}^{(k)}\,\mathcal{C}_k^b}_{\text{normalization closure}}$$
$$+ \underbrace{\sum_{k\in\{\text{SP},\text{VP}\}} \sum_{i=1}^{N_{\text{bins}}} \lambda_{\mathcal{C},i}^{(k)}\,\mathcal{C}_{k,i}^b}_{\text{shape-level closure}} .$$

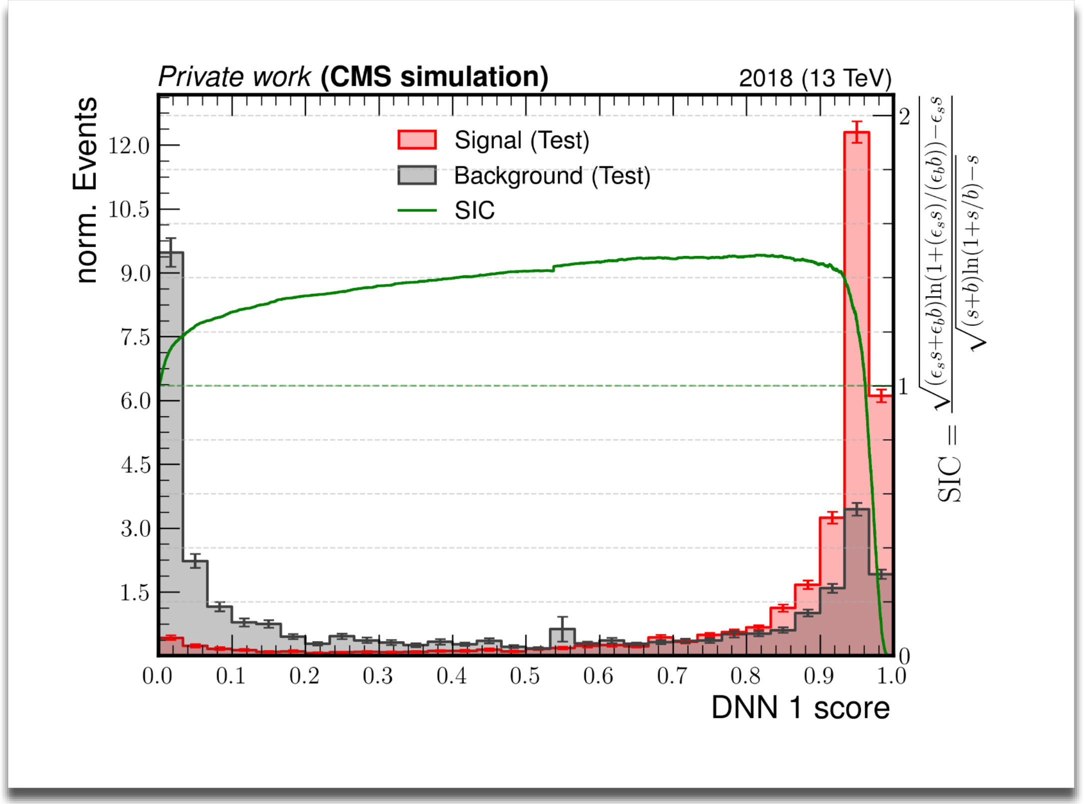
$$\mathcal{C}^b = \frac{\left|n_A^b n_D^b - n_B^b n_C^b\right|}{n_A^b n_D^b + n_B^b n_C^b}\,,$$

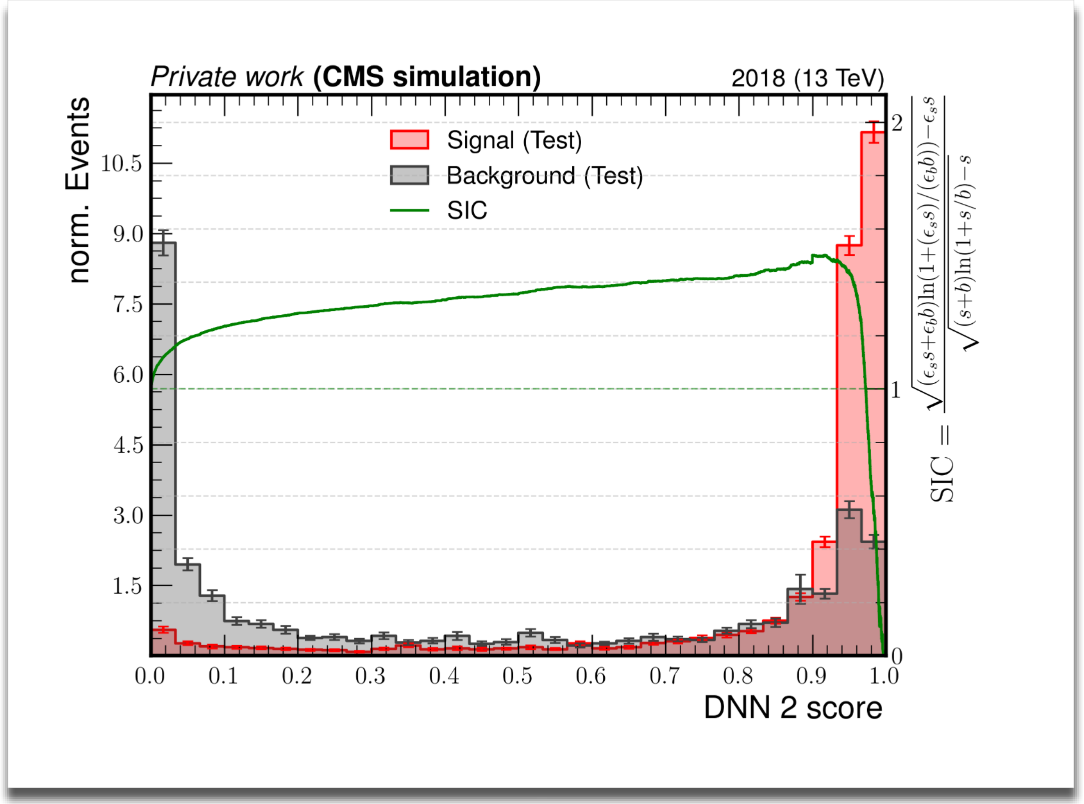
$$\mathcal{C}_i^b = \frac{\left|n_{A,i}^b n_{D,i}^b - n_{B,i}^b n_{C,i}^b\right|}{n_{A,i}^b n_{D,i}^b + n_{B,i}^b n_{C,i}^b}\,,$$

$$n_A^b = \sum_{i\in\mathcal{B}} G_\kappa(X_i - x_0)\,G_\kappa(y_0 - Y_i)\,w_i\,,$$
$$n_B^b = \sum_{i\in\mathcal{B}} G_\kappa(X_i - x_0)\,G_\kappa(Y_i - y_0)\,w_i\,,$$
$$n_C^b = \sum_{i\in\mathcal{B}} G_\kappa(x_0 - X_i)\,G_\kappa(y_0 - Y_i)\,w_i\,,$$
$$n_D^b = \sum_{i\in\mathcal{B}} G_\kappa(x_0 - X_i)\,G_\kappa(Y_i - y_0)\,w_i\,.$$

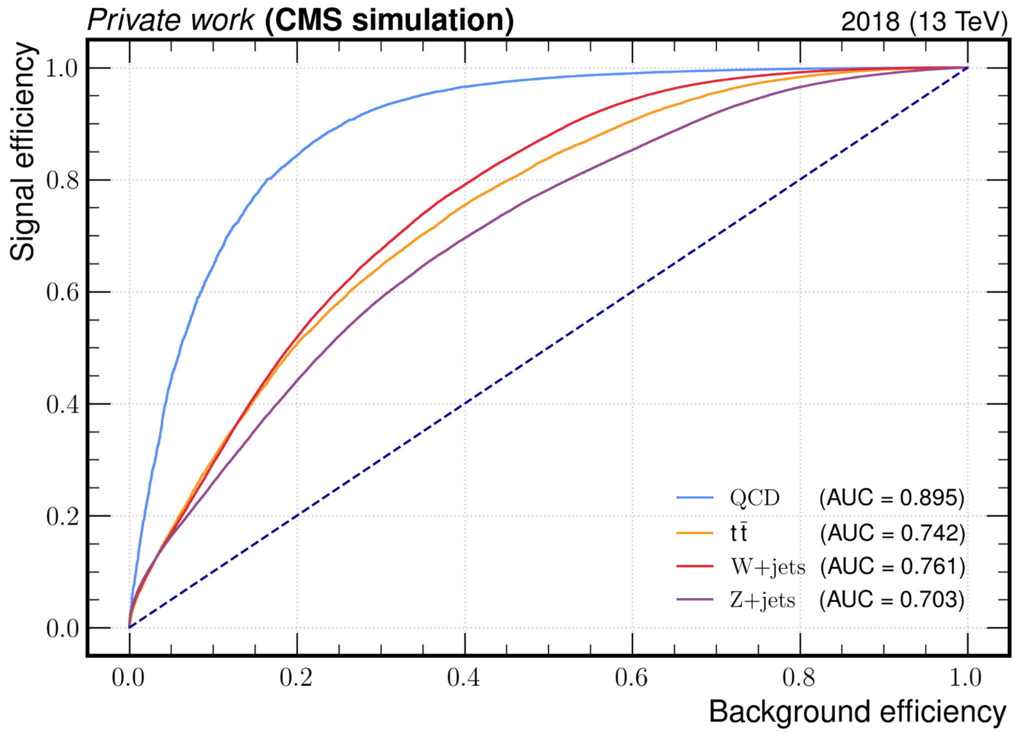
$$G_\kappa(u) = \sigma(\kappa u)$$

III) Score Distributions in the Validation Plane (VP)

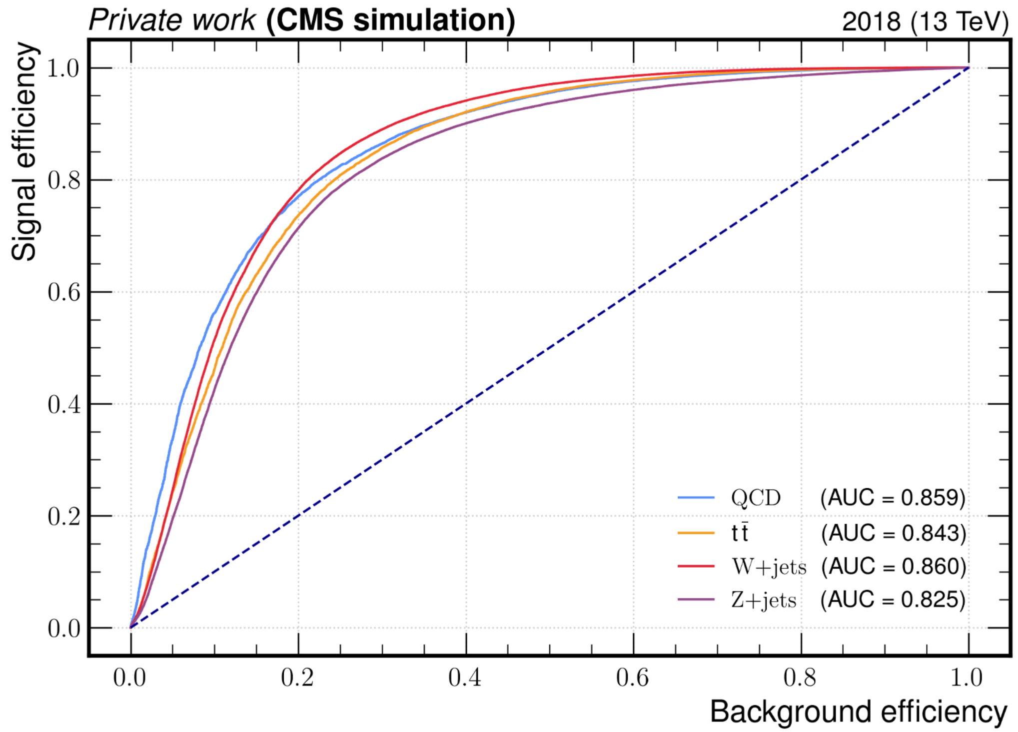




III) ROCs (SP)

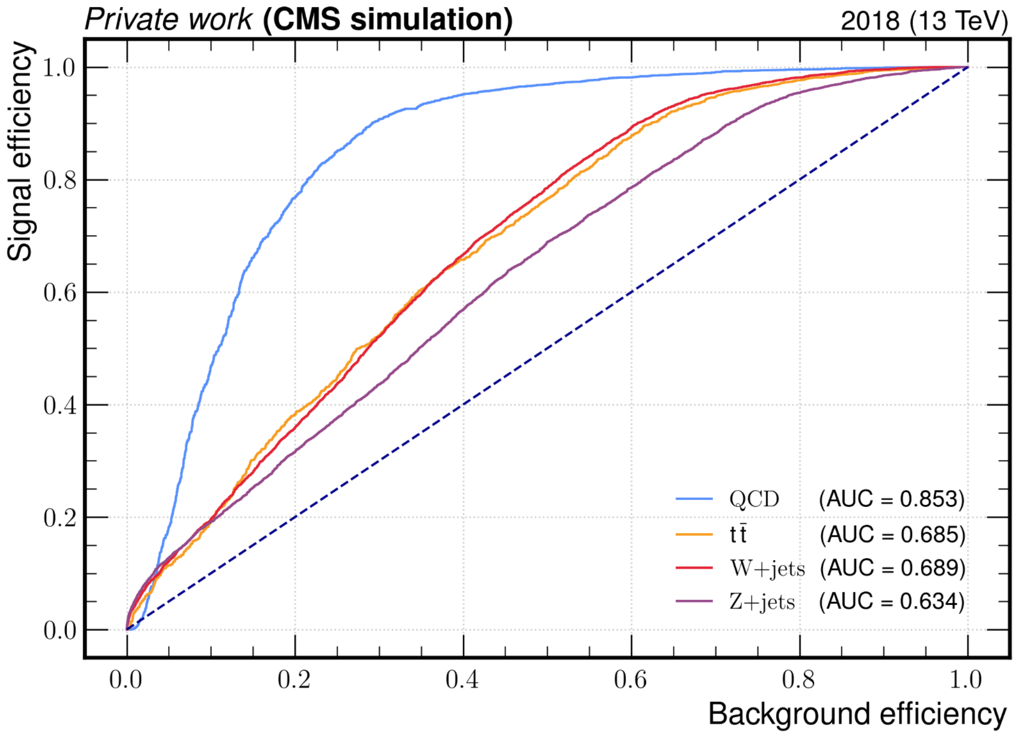


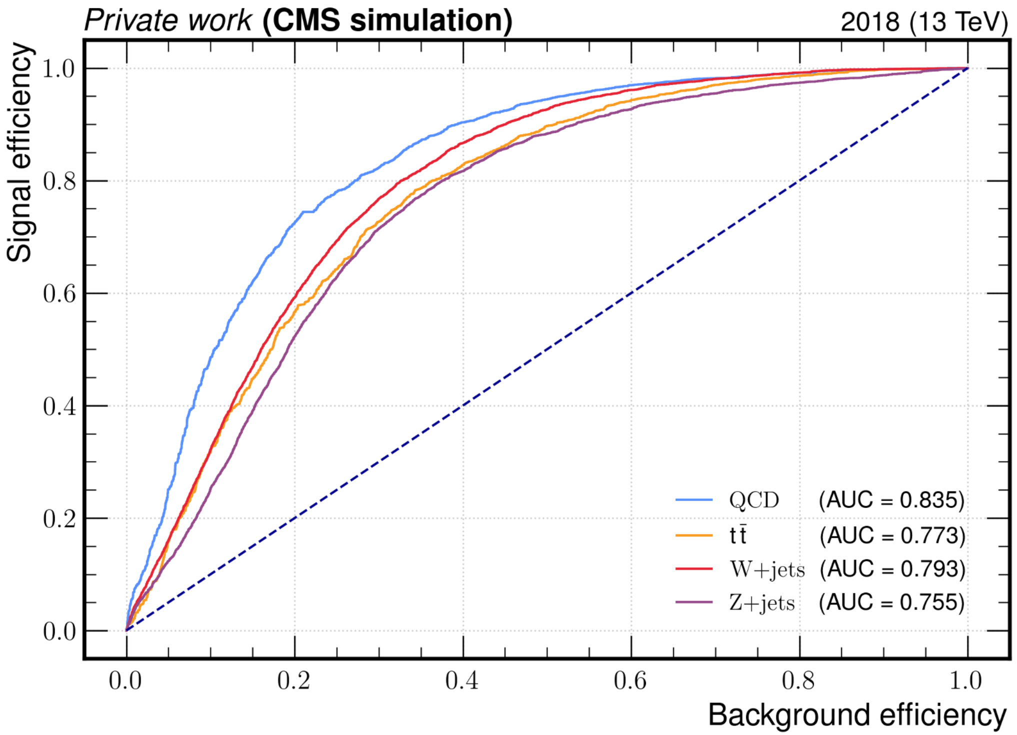
Private work (CMS simulation)
2018 (13 TeV)
Signal efficiency
Background efficiency
QCD (AUC = 0.895)
$t\bar{t}$ (AUC = 0.742)
W+jets (AUC = 0.761)
Z+jets (AUC = 0.703)



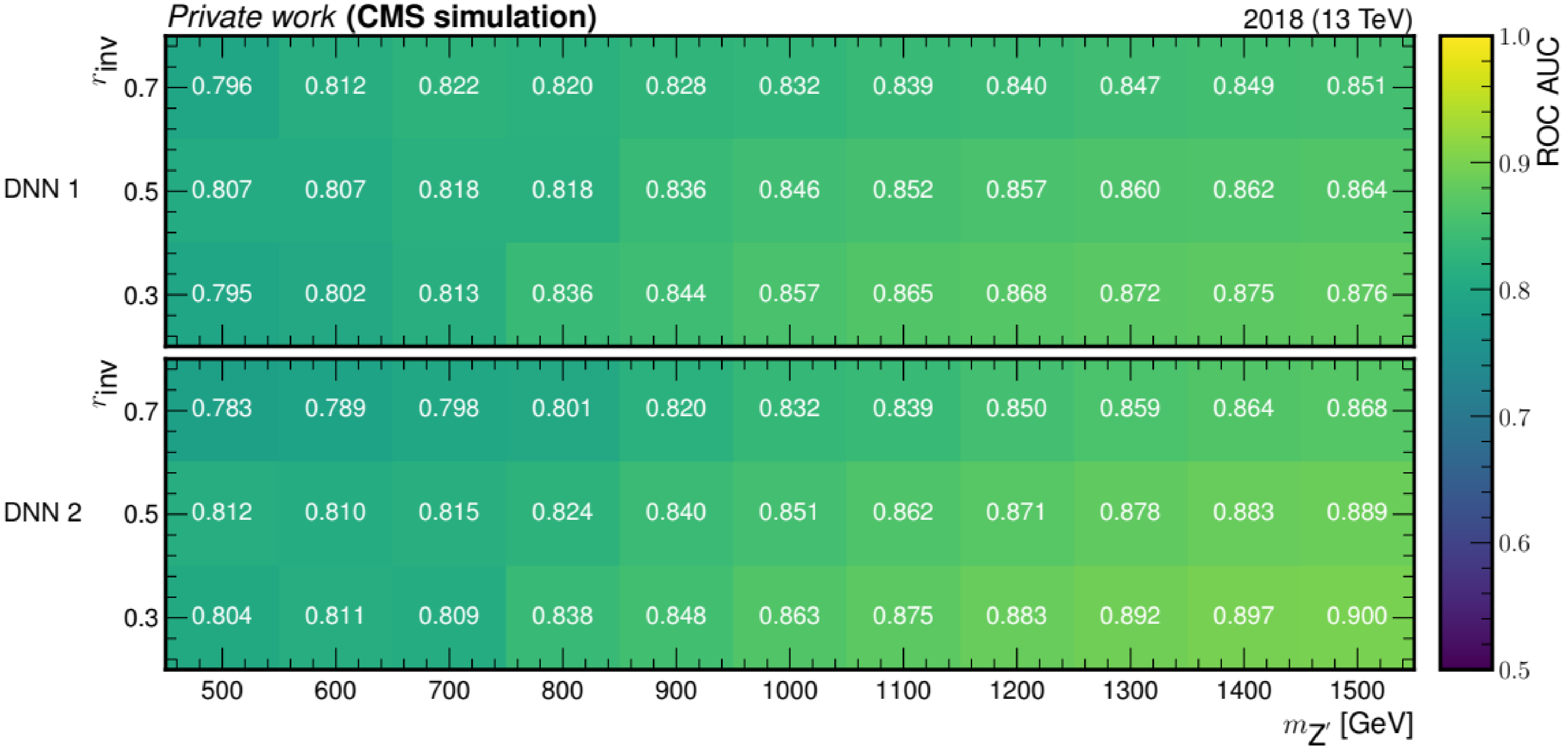
Private work (CMS simulation)
2018 (13 TeV)
Signal efficiency
Background efficiency
QCD (AUC = 0.859)
$t\bar{t}$ (AUC = 0.843)
W+jets (AUC = 0.860)
Z+jets (AUC = 0.825)

III) ROCs (VP)

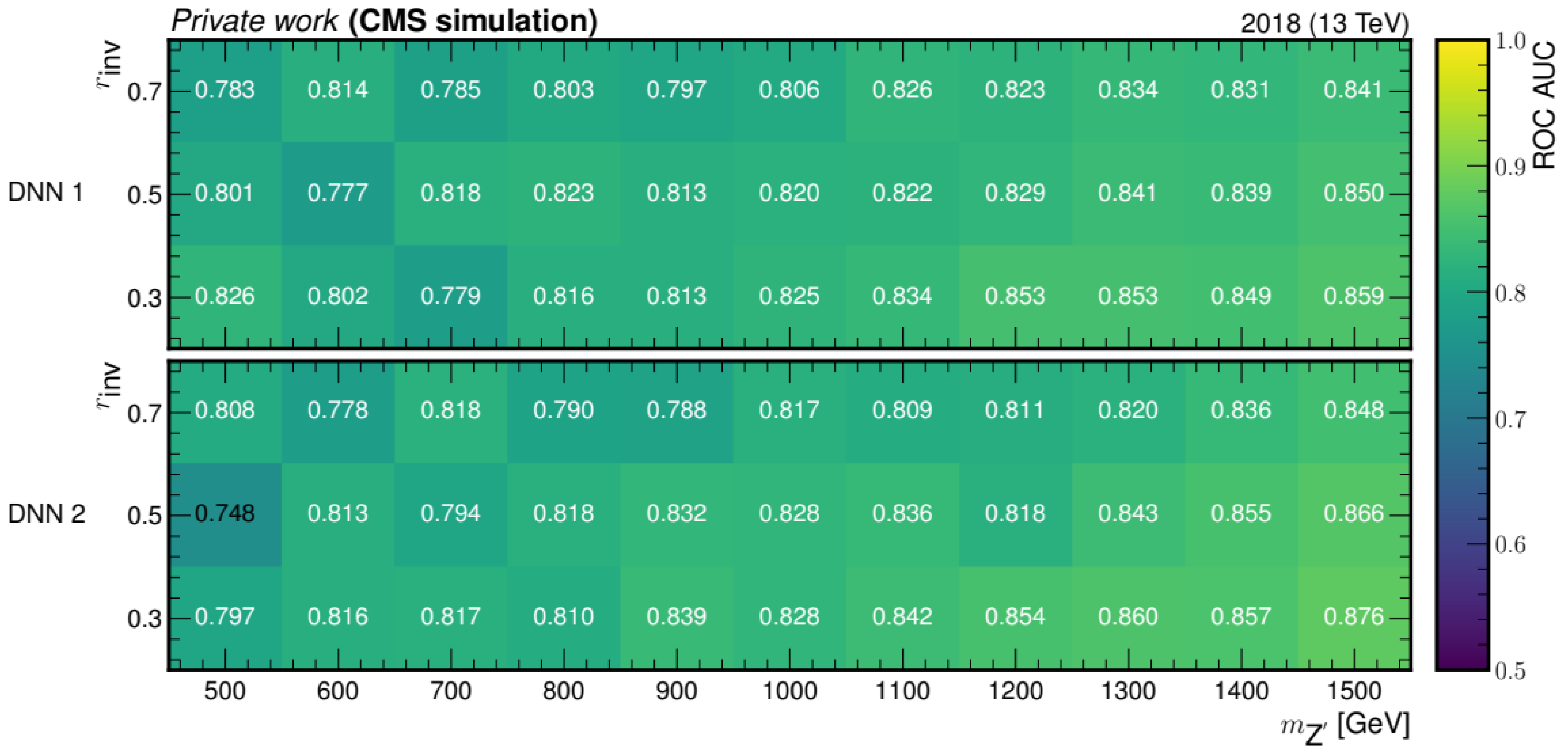

Private work (CMS simulation)
2018 (13 TeV)
Signal efficiency
Background efficiency
0.0
0.2
0.4
0.6
0.8
1.0
QCD (AUC = 0.853)
$t\bar{t}$ (AUC = 0.685)
W+jets (AUC = 0.689)
Z+jets (AUC = 0.634)


Private work (CMS simulation)
2018 (13 TeV)
Signal efficiency
Background efficiency
0.0
0.2
0.4
0.6
0.8
1.0
QCD (AUC = 0.835)
$t\bar{t}$ (AUC = 0.773)
W+jets (AUC = 0.793)
Z+jets (AUC = 0.755)

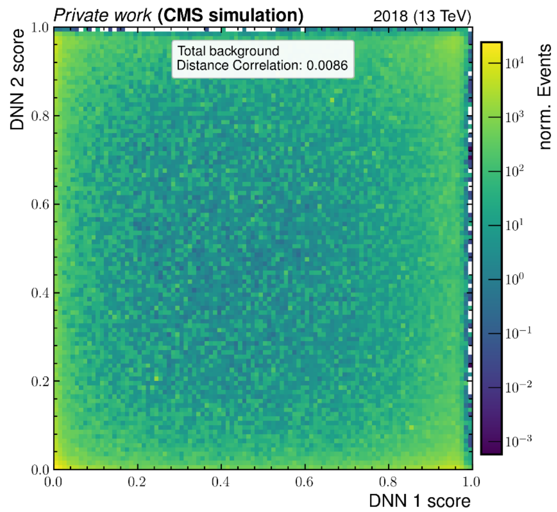
III) ROC AUCs (SP)

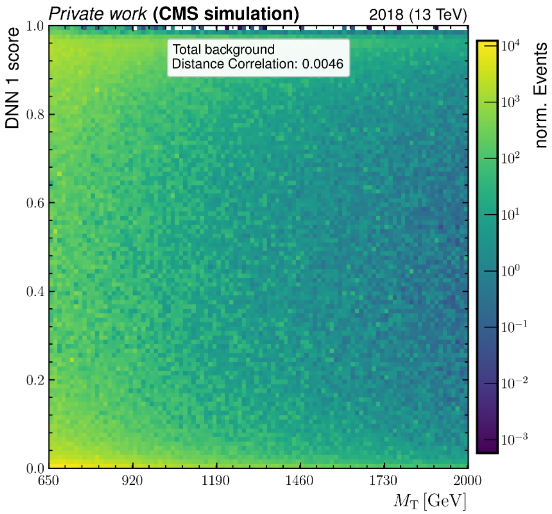


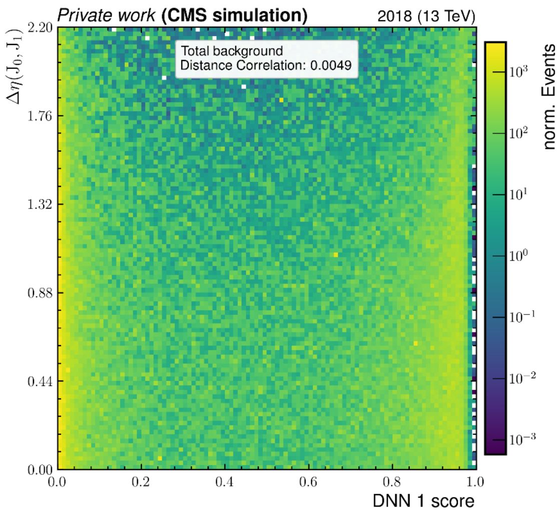
III) ROC AUCs (VP)

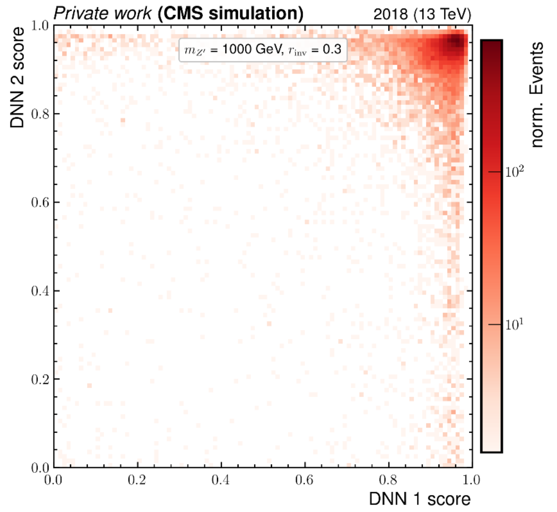


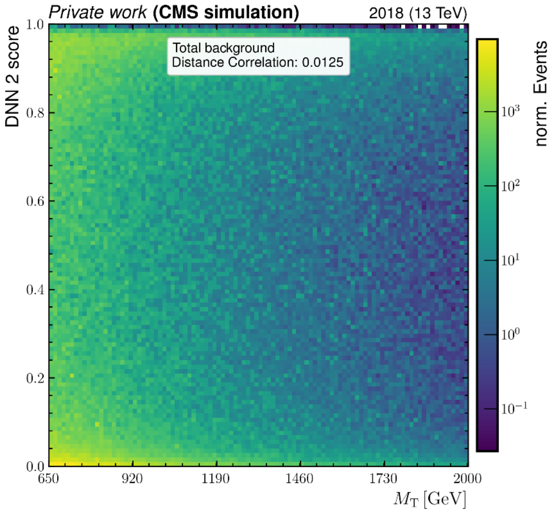
III) 2D Distributions & DisCo

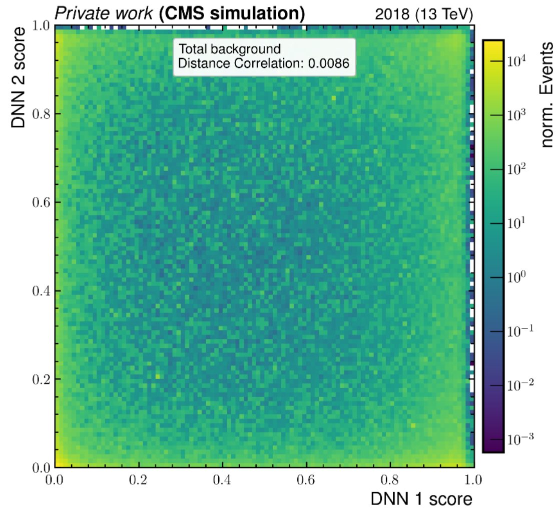




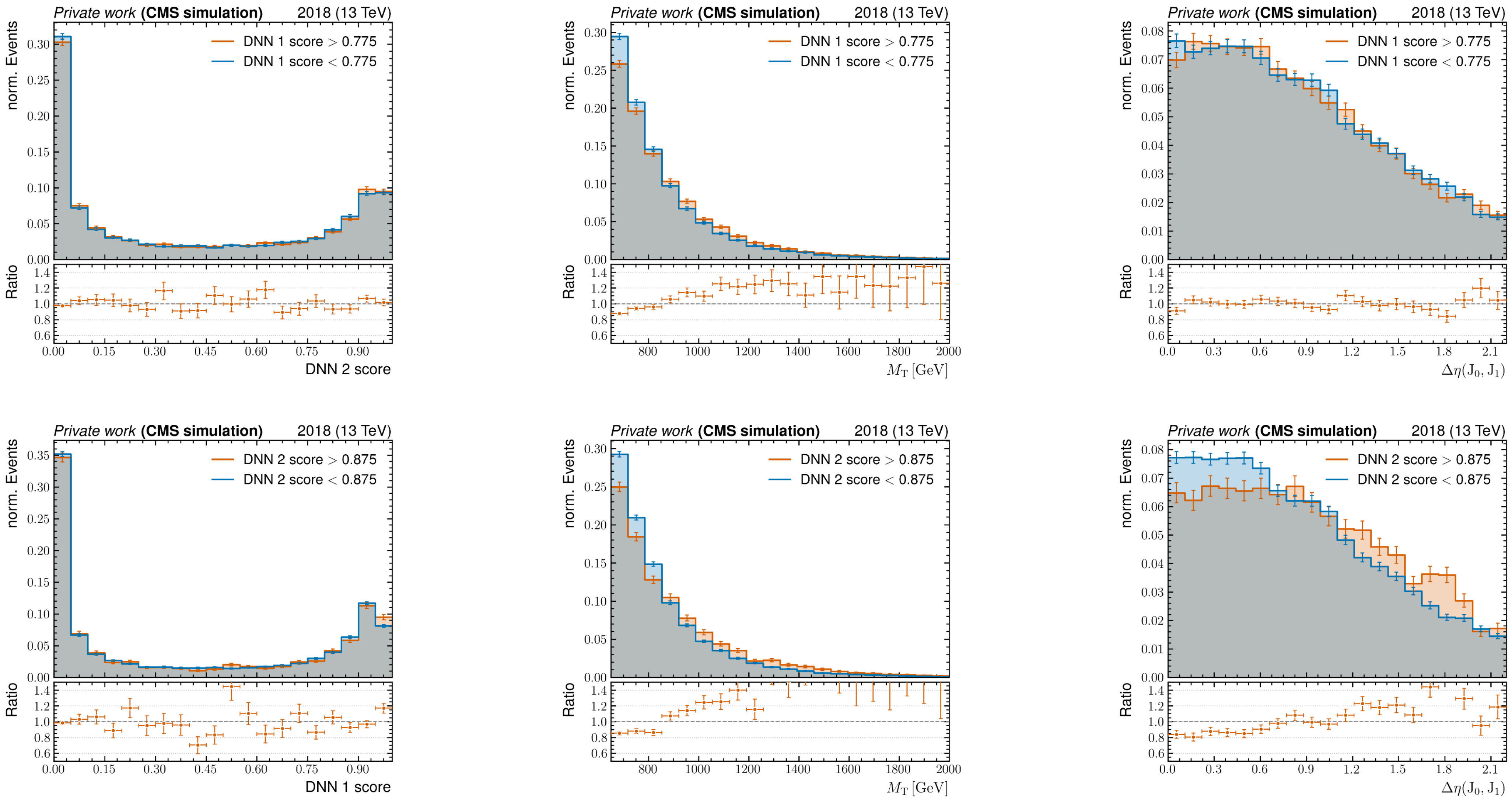








III) 1D Impact Plots



III) Boundary Optimization – FOM Details

Karlsruhe Institute of Technology

Metric	Rule
Mean mod. binned Asimov significance $\langle Z^{\mathcal{A}}_{\not{C}} \rangle$	maximize
SR χ^2 test p-value	$\geq 90\%$ and on a local plateau
Absolute fractional non-closure $\lvert\delta_{\not{C}}\rvert$	$\leq 5\%$
$\lvert$Pull$\rvert$ (normalization)	≤ 2.0
Max. normalized signal contamination $r_{\max}$	≤ 0.35
WP bounds	$X, Y \in [0.1, 0.9]$

$$\left\langle Z^{\mathcal{A}}_{\not{C}} \right\rangle = \frac{1}{N_\sigma} \sum_{k=1}^{N_\sigma} Z^{\mathcal{A}}_{\not{C},k}$$

$$Z^{\mathcal{A}}_{\not{C},k} = \left[\sum_{i=1}^{N_{\text{bins}}} \left(Z^{\mathcal{A}}_{\not{C},k,i} \right)^2 \right]^{1/2}$$

$$\delta_{\not{C}} \equiv \frac{\hat{n}^b_B}{n^b_B} - 1$$

$$|\text{Pull}| \equiv \frac{\left| \hat{n}^b_B - n^{b,\text{MC}}_B \right|}{\sqrt{\sigma^2\left(\hat{n}^b_B\right) + \sigma^2\left(n^{b,\text{MC}}_B\right)}}$$

$$\tilde{s}_{k,i} \equiv n^s_{B,k,i} \frac{\sum_{R\in\{A,B,C,D\}} n^s_{R,l}}{\sum_{R\in\{A,B,C,D\}} n^s_{R,k}}$$

$$\left\langle \tilde{Z}_{\not{C}} \right\rangle = \frac{1}{N_\sigma} \sum_{k=1}^{N_\sigma} \frac{\tilde{s}_k}{\sqrt{b + \tilde{s}_k + |\delta_{\not{C}}|\, b}}$$

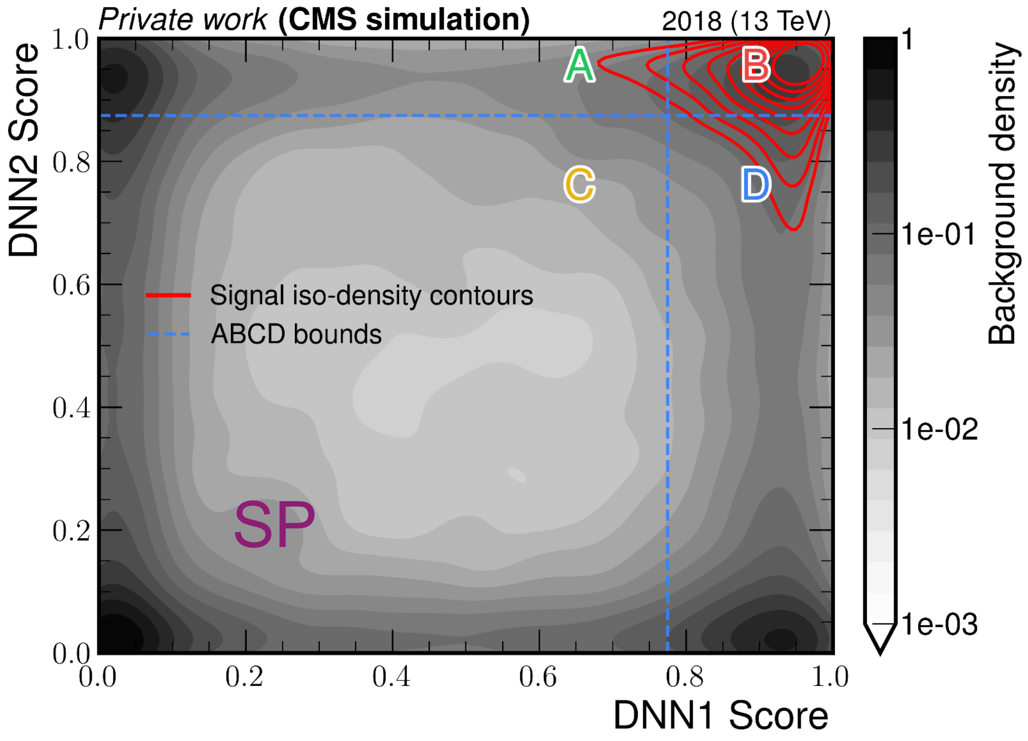
$$Z^{\mathcal{A}}_{\not{C},k,i} = \left\{ 2 \left[\left(\tilde{s}_{k,i} + b_i\right) \ln\left(\frac{\left(\tilde{s}_{k,i} + b_i\right)\left(b_i + \sigma^2_{b,i}\right)}{b_i^2 + \left(\tilde{s}_{k,i} + b_i\right)\sigma^2_{b,i}} \right) - \frac{b_i^2}{\sigma^2_{b,i}} \ln\left(1 + \frac{\sigma^2_{b,i}\, \tilde{s}_{k,i}}{b_i\left(b_i + \sigma^2_{b,i}\right)} \right) \right] \right\}^{1/2}$$

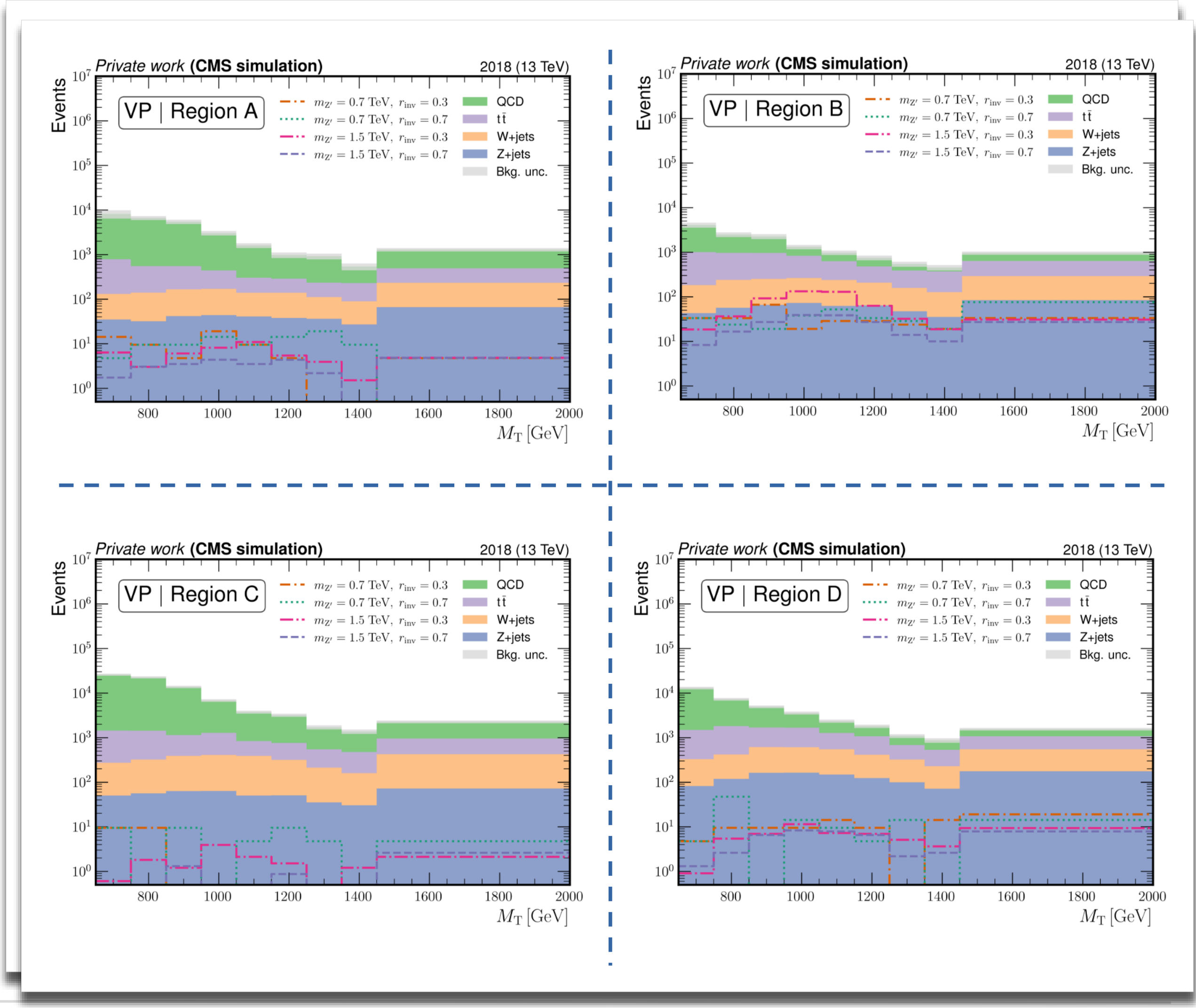
https://www.pp.rhul.ac.uk/~cowan/stat/medsig/medsigNote.pdf

III) Boundary Optimization – Applying the Working Point

Karlsruhe Institute of Technology

And similarly for the VP...



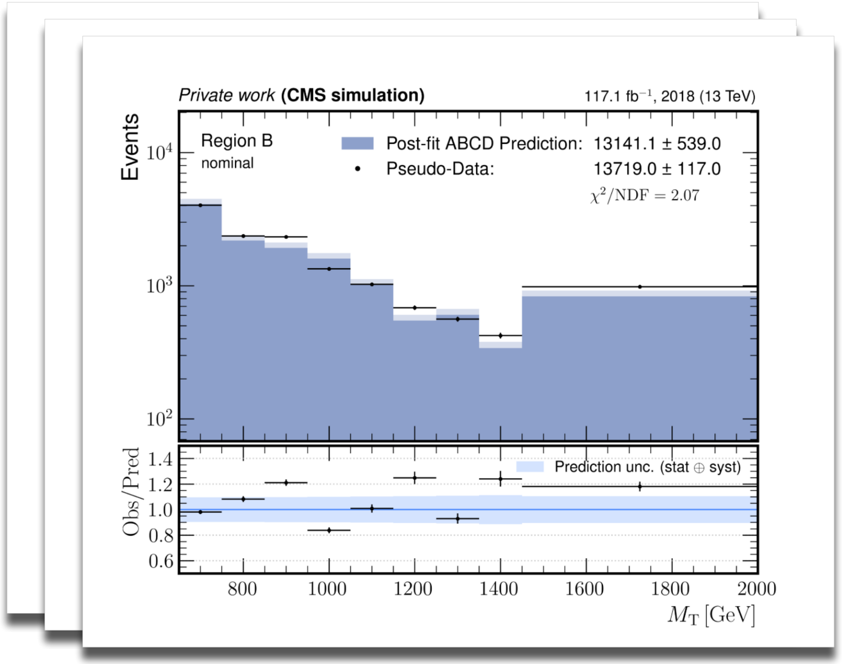


III) Expected Background Yields at optimized WP

Process	Plane	Region A	Region B	Region C	Region D	Total
QCD	SP	102541	25086	376254	108167	612048
	VP	24566	7492	72132	24993	129183
$t\bar{t}$	SP	18187	26432	71890	78808	195317
	VP	2545	4299	5949	7246	20039
W+jets	SP	7600	7996	32803	29686	78085
	VP	956	1425	2420	2876	7677
Z+jets	SP	2257	2621	5406	9736	20020
	VP	359	531	472	1145	2507
Total	SP	130585	62135	486353	226397	905470
	VP	28426	13747	80973	36260	159406

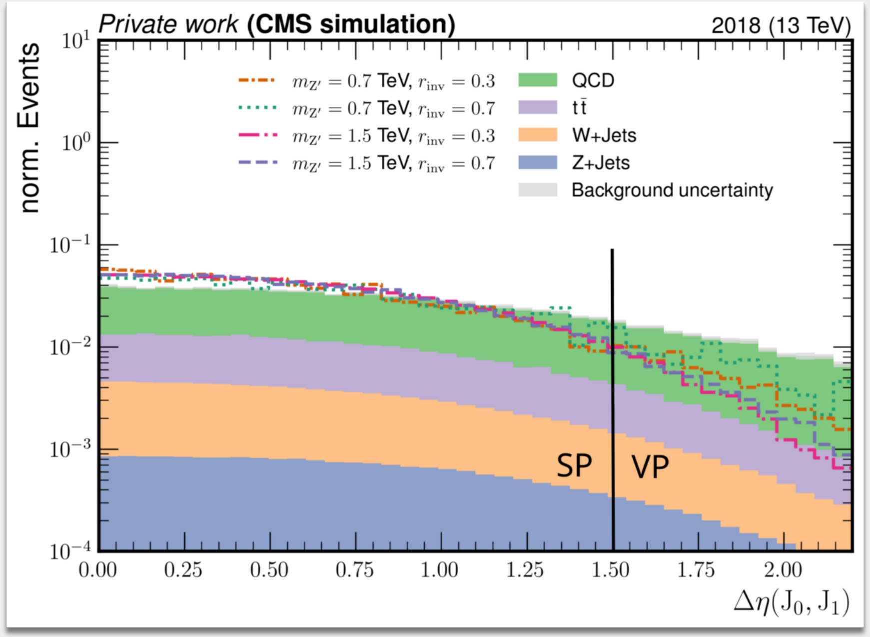
III) Validation

- To demonstrate: stable, unbiased background prediction at chosen WP before unblinding
- Key principles - Validate:
  - ABCD closure
  - Stability against variations of ABCD boundaries
  - Robustness to signal contamination
- Progress from pseudo-data (SP, VP) to data (VP)



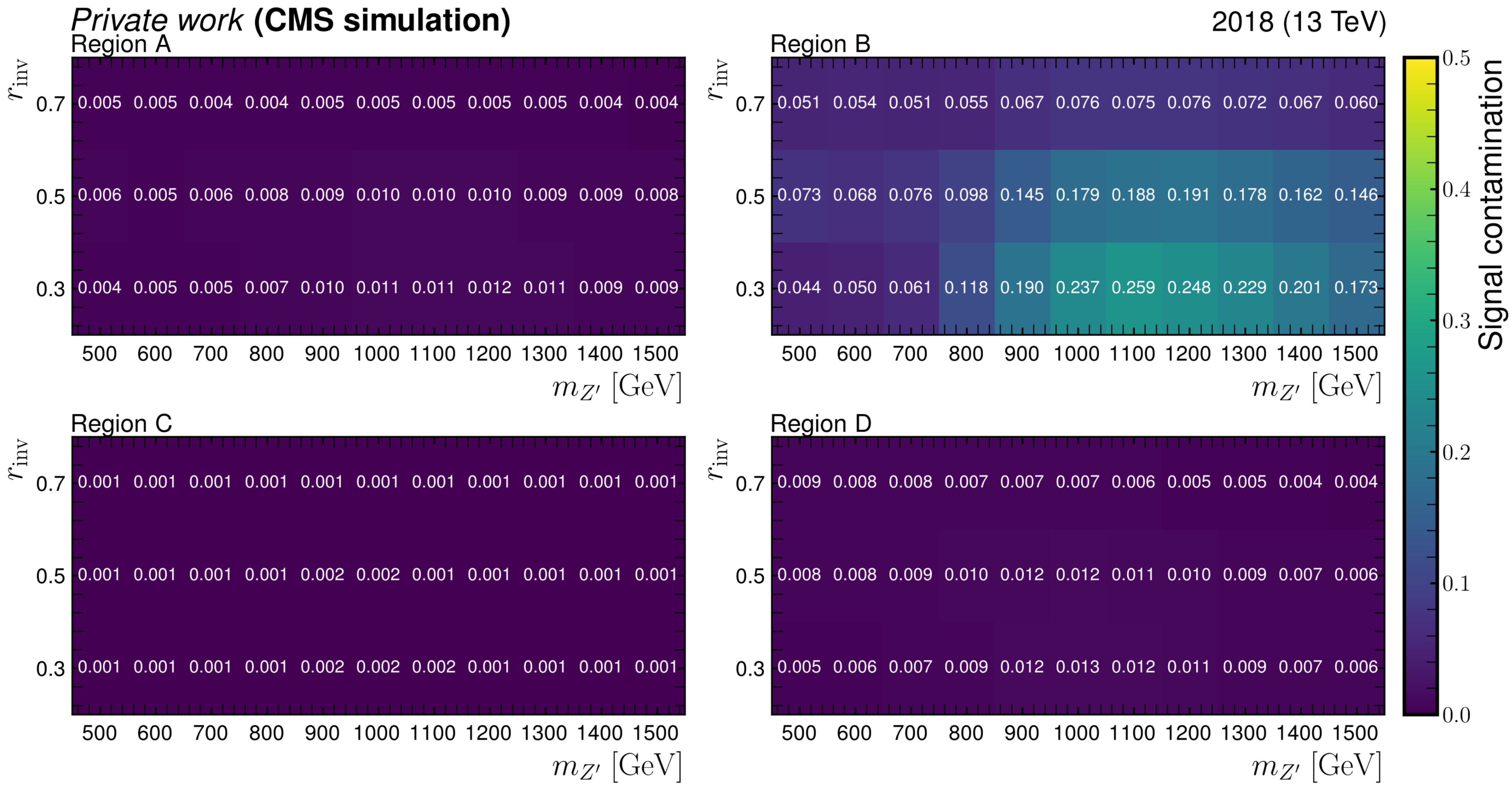
?

Reminder:

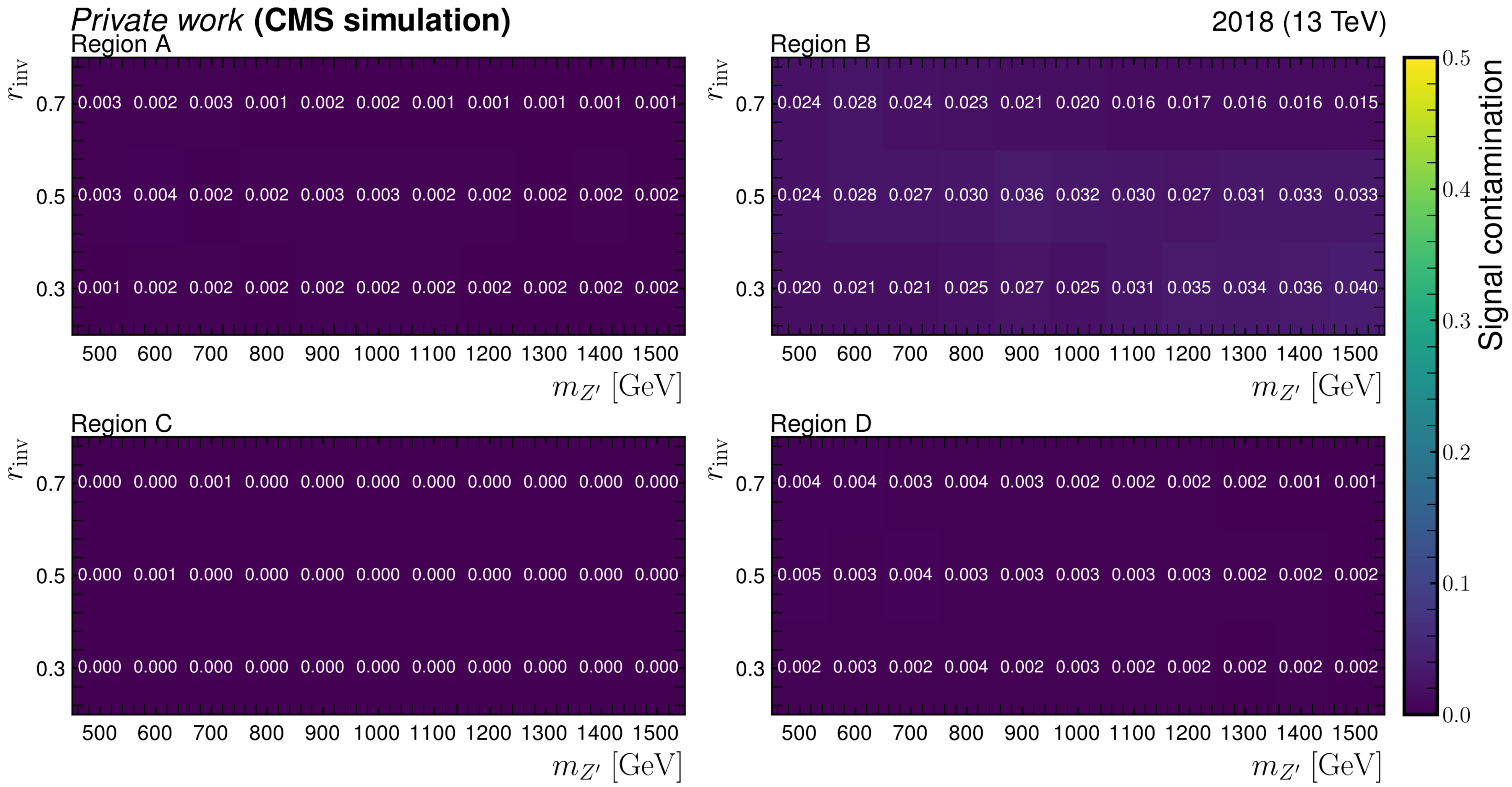


(Distributions after applying preselection)

III) Signal contamination panels (SP)

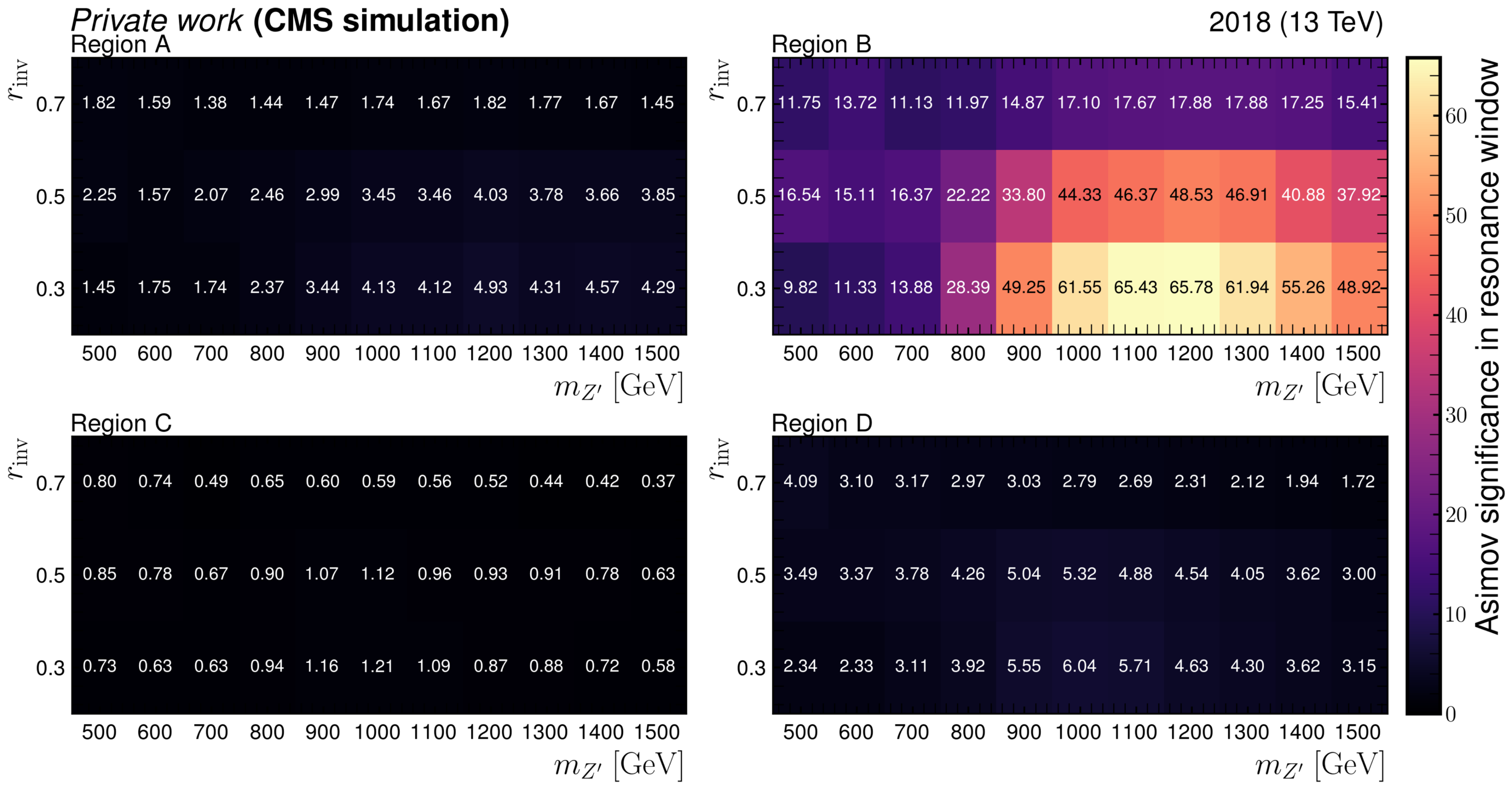


III) Signal contamination panels (VP)

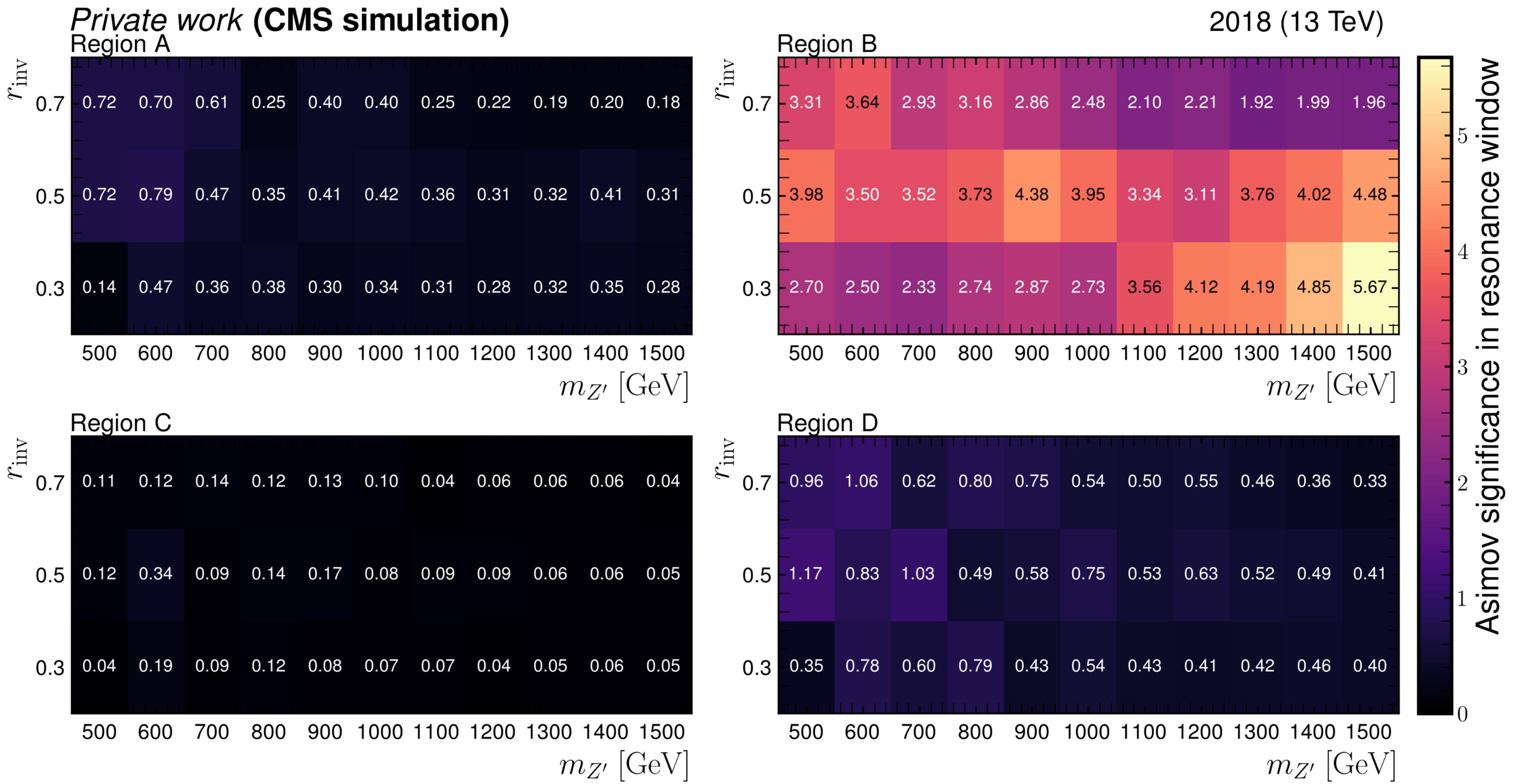


III) Signal significance panels (SP)

Karlsruhe Institute of Technology

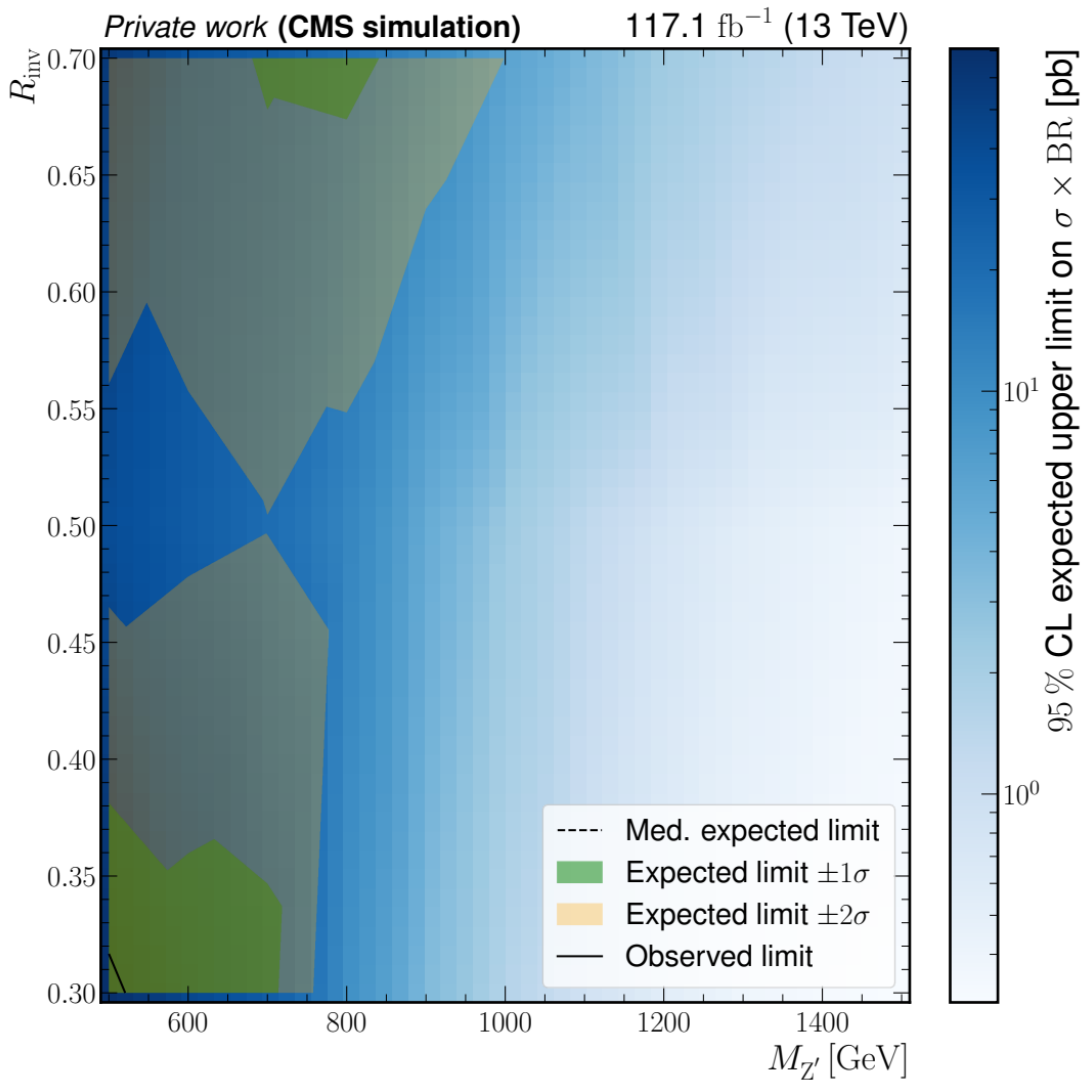


III) Signal significance panels (VP)



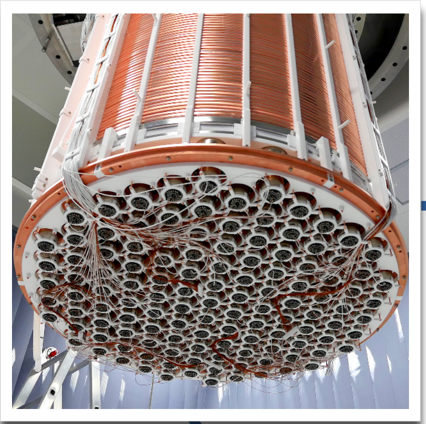
IV) Statistical Inference: Expected Limits (2D)

- Performed with CMS Combine via CLs method
- Using the profile-likelihood test statistic & the asymptotic approximation
- Combined S+B fit across regions {A,B,C,D} per bin
- Include per-bin non-closure systematics for background
- Signal systematics to be added



DM Search Strategies

Direct

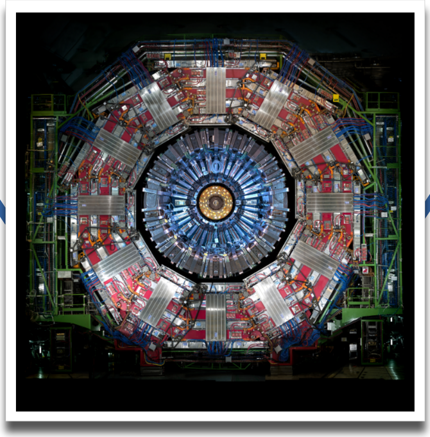


XENON Collaboration

Indirect

NASA

Collider



CMS Collaboration

A Supervised Machine-Learning Search for Semi-Visible Jets in CMS Run-2 Scouting Data

Aimar Aguado Berasaluce, Cesare Tiziano Cazzaniga, Annapaola de Cosa, Marcel Gaisdörfer, Rebecca Natalia Hamp, Celeste Holm, **Jonas Janik**, Markus Klute, Benedikt Maier, Kevin Pedro, Brendan Regnery, Roberto Seidita