DRAPES: Diffusion for weakly supervised searches

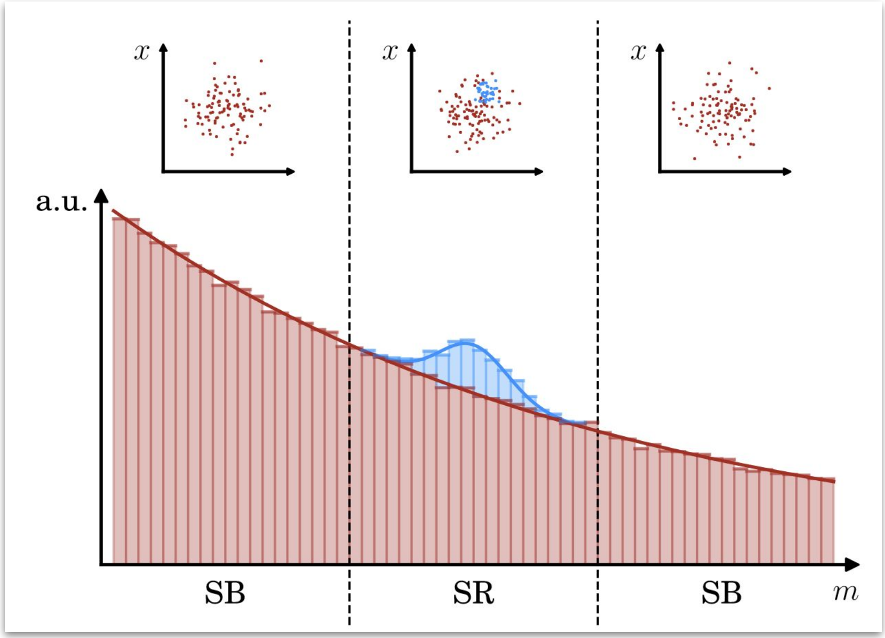
ML4Jets, 2023

Debajyoti Sengupta, **Matthew Leigh**, Johnny Raine, Sam Klein, Tobias Golling

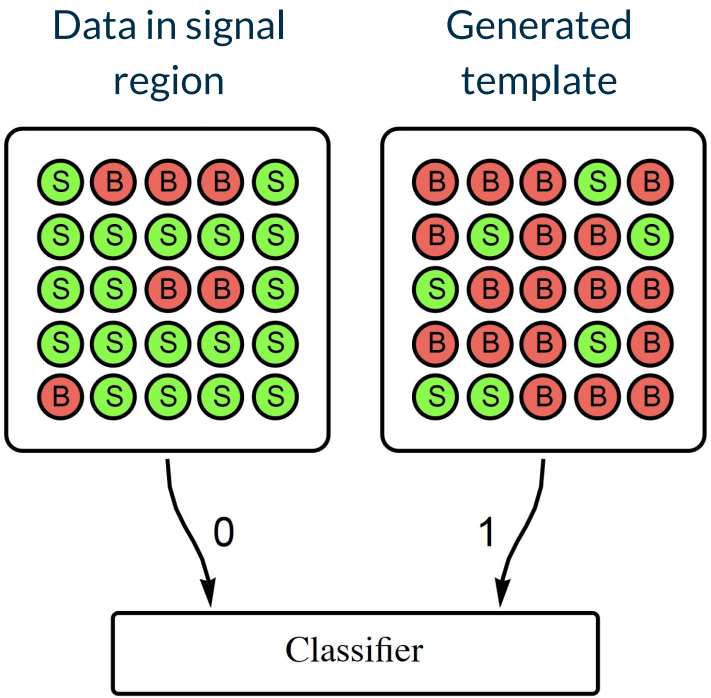
UNIVERSITÉ DE GENÈVE

Template Building

- **CATHODE**:
  - Purely data driven
  - Train conditional model in the sidebands
    - Conditional on mjj
  - Generate template in the signal region



CWoLa



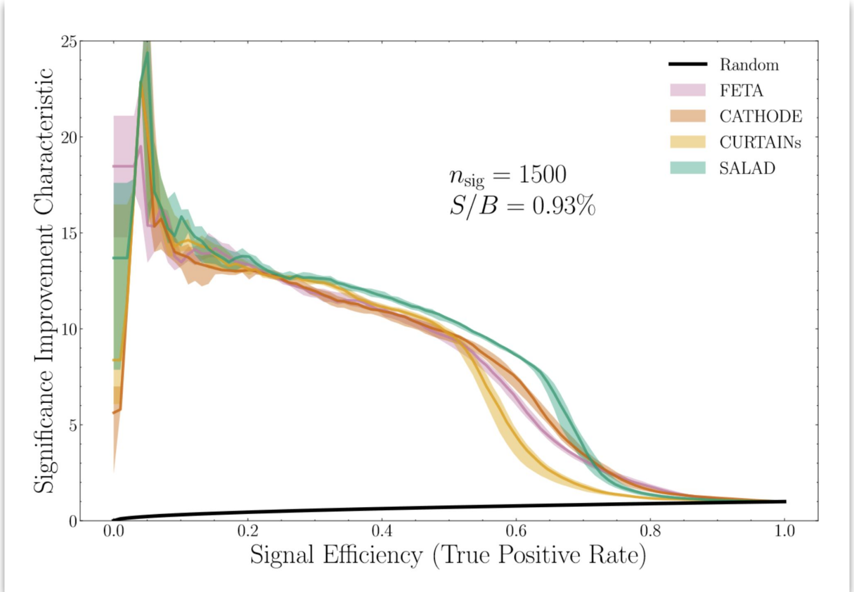
Data in signal region
Generated template
S B B B S
S S S S S
S S B B S
S S S S S
B S S S S
B B B S B
B S B B S
S B B B B
B B B S B
S S B B B
0
1
Classifier

Weakly Supervised Regime

- LHCO RnD dataset
  - Background: QCD dijets
  - Signal: W’ → X(qq) Y(qq)

- Dijet system described by a 5-vector

$$m_{J_1},\ \Delta m_J = m_{J_1} - m_{J_2},\ \tau_{21}^{J_1},\ \tau_{21}^{J_2},\ \Delta R_{JJ} = \sqrt{\Delta\eta^2 + \Delta\phi^2}$$



from : 2307.11157

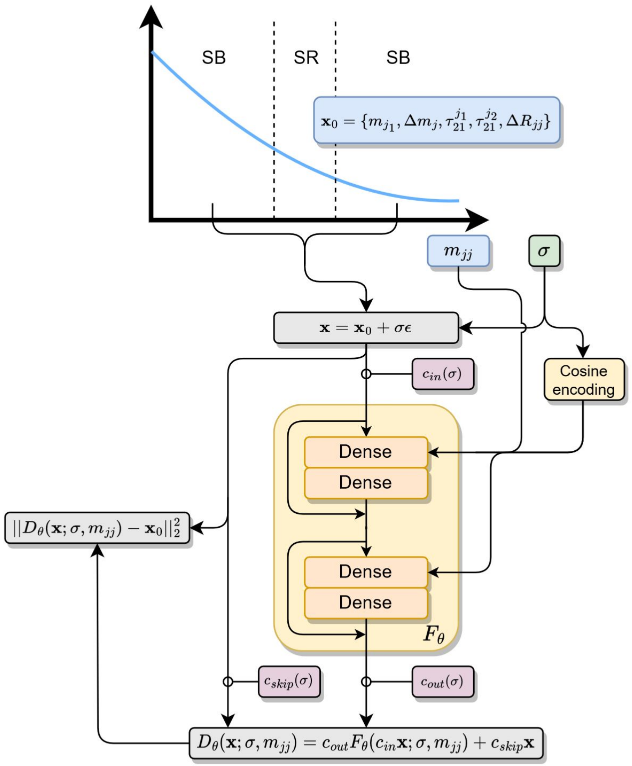
Task

- All ML methods so far use normalising flows
- Lately we have seen great success with diffusion models
- Lets try them for weakly supervised anomaly detection!
    - Benchmark using LHCO

Drapes Ø: Training

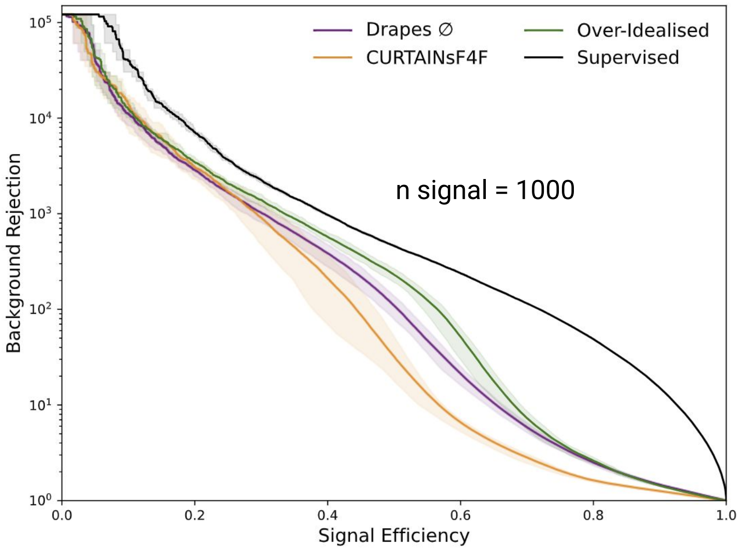
- Template building with diffusion on features
    - Train on sidebands
    - Generate in signal region
    - Analogous to CATHODE

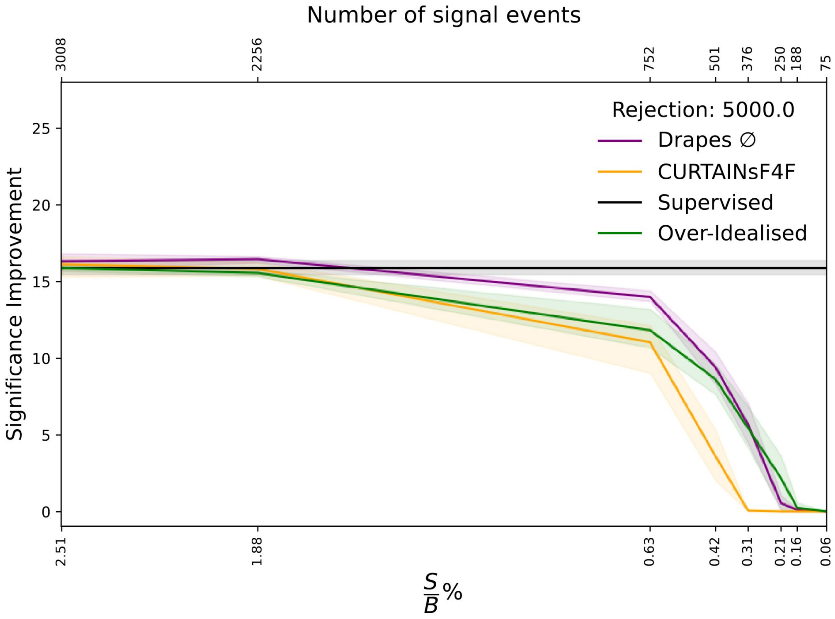
- Step one - use same features as all the others



Drapes Ø: Performance

- Diffusion models achieve state of the art!





Comparing to CURTAINsF4F with max width sidebands (2305.04646), same conclusions comparing to CATHODE

What else?

- Diffusion improves on SotA performance with same inputs!
- But also: Diffusion opens up further possibilities

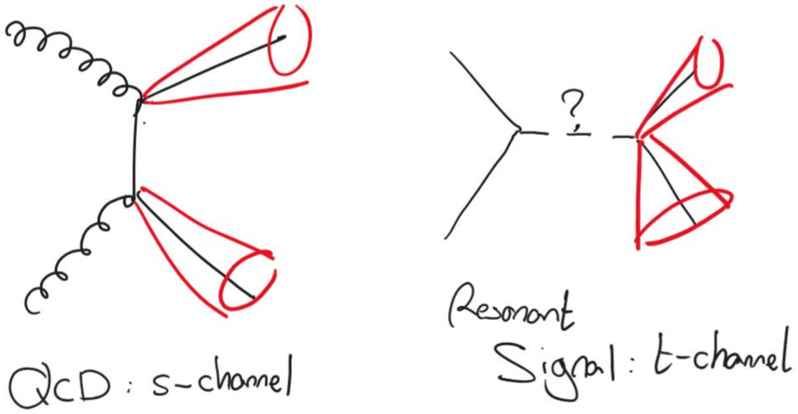
1. High quality generation of point cloud data
2. Partial generation

Anomaly detection with point cloud data

Drapes Ø for constituent level

- Instead of the high level features, train a diffusion model to generate the jet point clouds
  - No longer need to choose "optimal" set

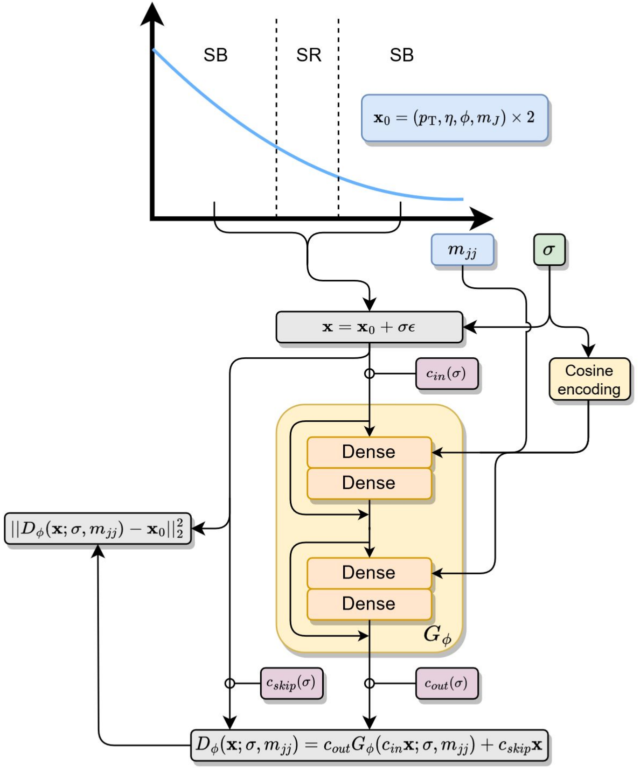
- Use PC-Droid model to conditionally generate jets
  - Can diffuse each jet independently
  - Don't expect correlated substructure in QCD
  - But additionally need to model jet kinematics | mjj



See also: Cedric's talk later today on Full Phase Space Anomaly Detection https://arxiv.org/abs/2307.06836

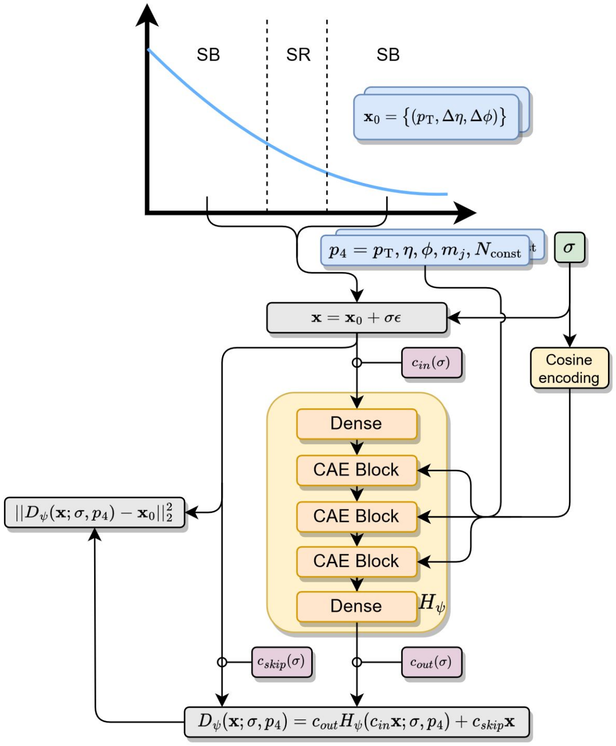
Generation Chain

- Model 1
  - Generate **jet kinematics** given **mjj**



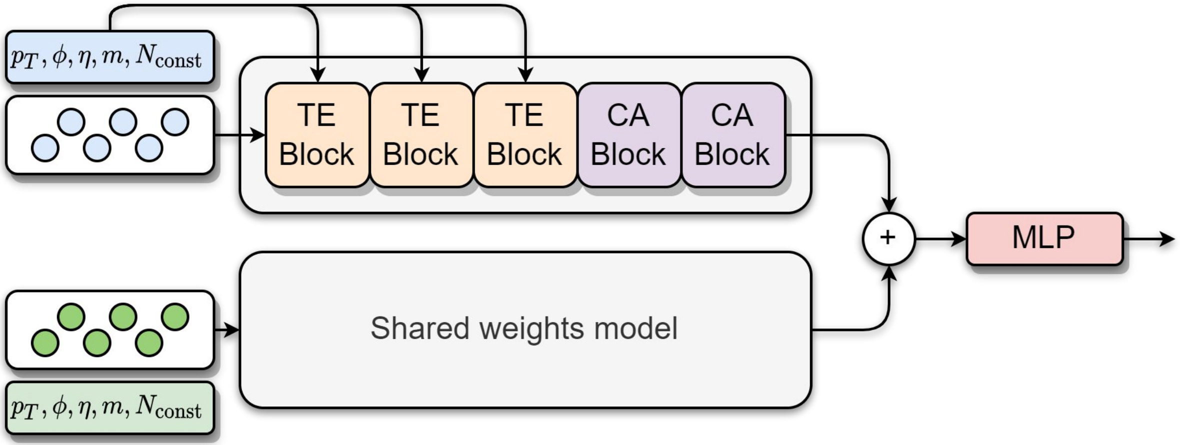
Generation Chain

- Model 2
  - Generate **point cloud** given **kinematics**



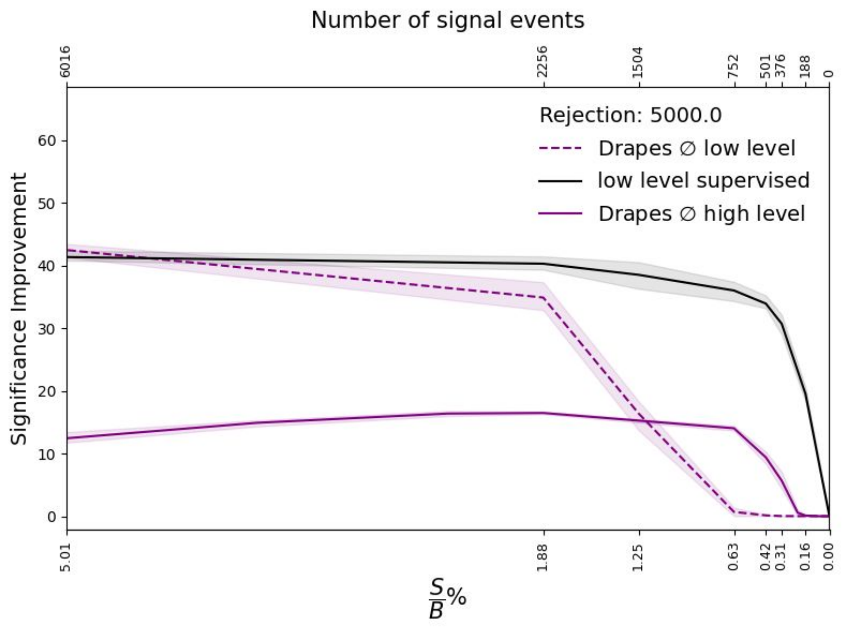
Discriminator used for CWoLa

- The two jets are processed by the same network.
- The outputs are added and passed through MLP
    - Permutation invariance within jets
    - Permutation invariance between jets



Drapes for constituent level

- Massive boost in performance for S/B > 1%:
  - CWoLa training struggles in higher dimensions
- High level features still performant for lower signal strengths
- Same behaviour observed in idealised setting



Anomaly detection with partial diffusion

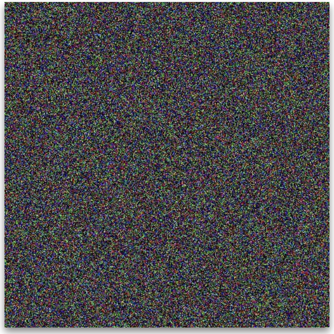
Partial Diffusion

Forward Process (add noise, no learning involved)

Sample data

t=0

Noise



t=1

Reverse Process (requires neural net)

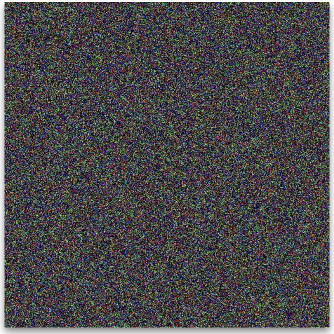
Partial Diffusion

Sample data

t=0

t = strength

Noise



t=1

Partial Diffusion

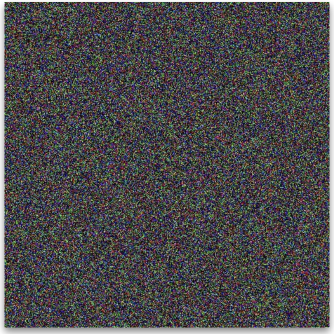
Sample data

t=0

“picasso painting”

t = strength

Noise

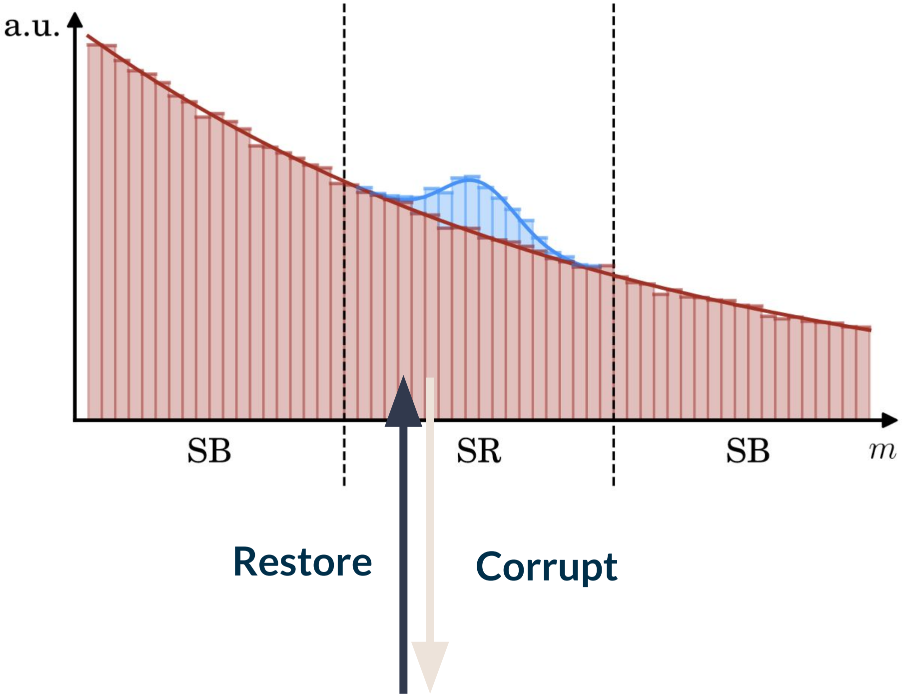


t=1

Drapes

- **Modify** not generate

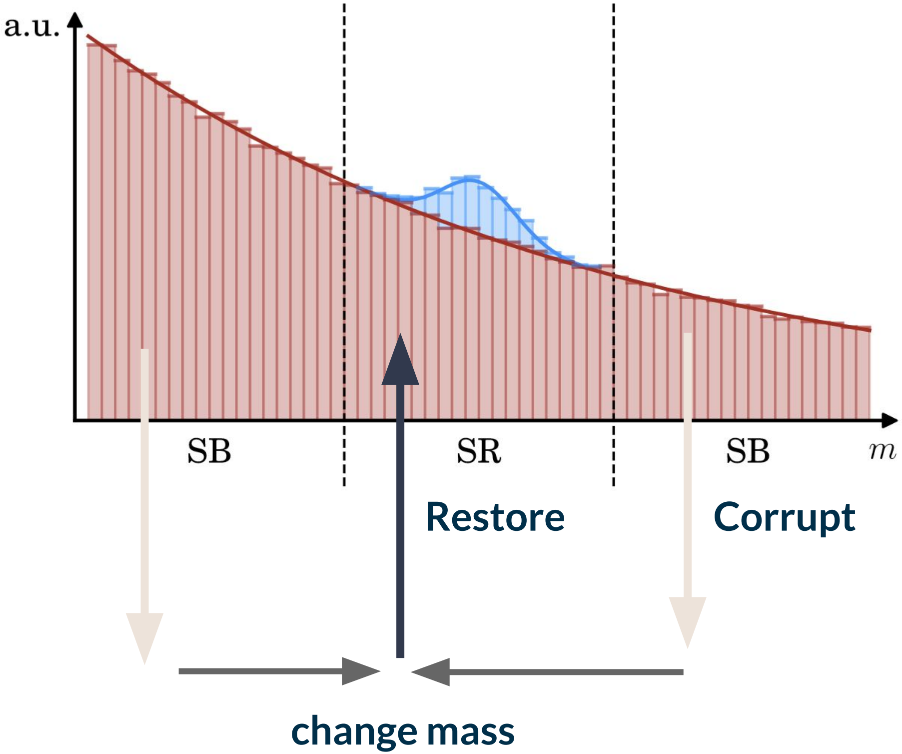
- Where do we modify **our data from**?
    - <u>**DRAPES SR**</u>
    - From the signal region
    - Should make signal samples less “signally”



Drapes

- **Modify** not generate

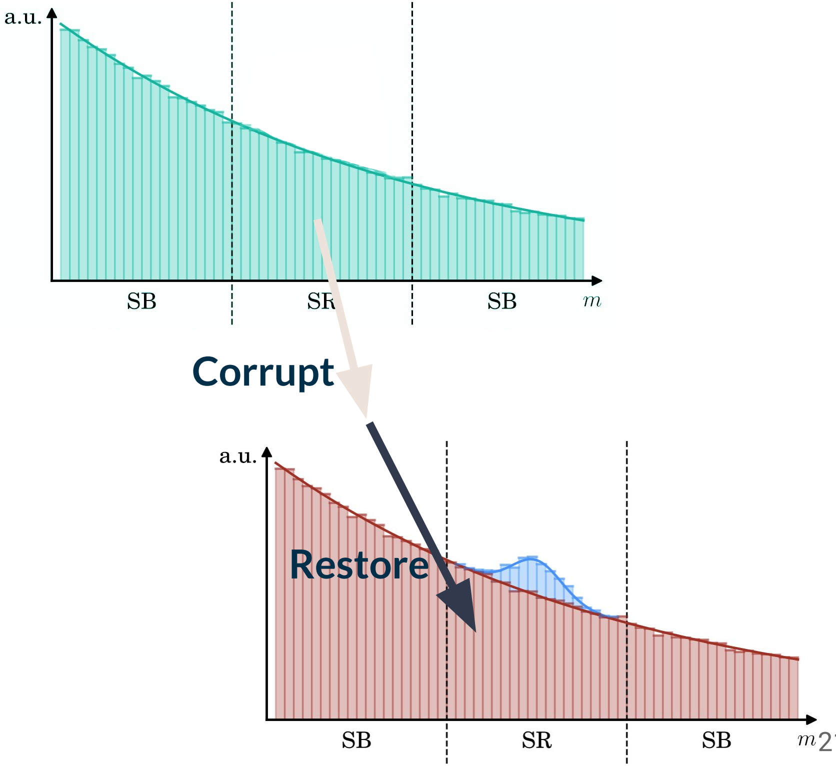
- Where do we modify **our data from**?
    - <u>**DRAPES SB**</u>
    - From the sideband
    - Give sample new mass
    - Similar to CURTAINS



Drapes

- **Modify** not generate

- Where do we modify **our data from**?
    - **DRAPES MC**
    - From the another MC template
    - Change sample generation
    - Similar to FETA



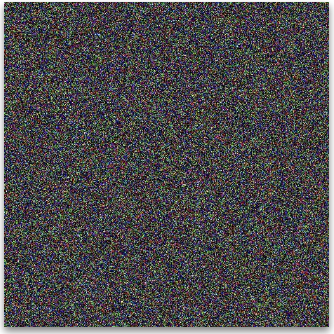
Partial Diffusion

Sample data

t=0

t = strength

Noise

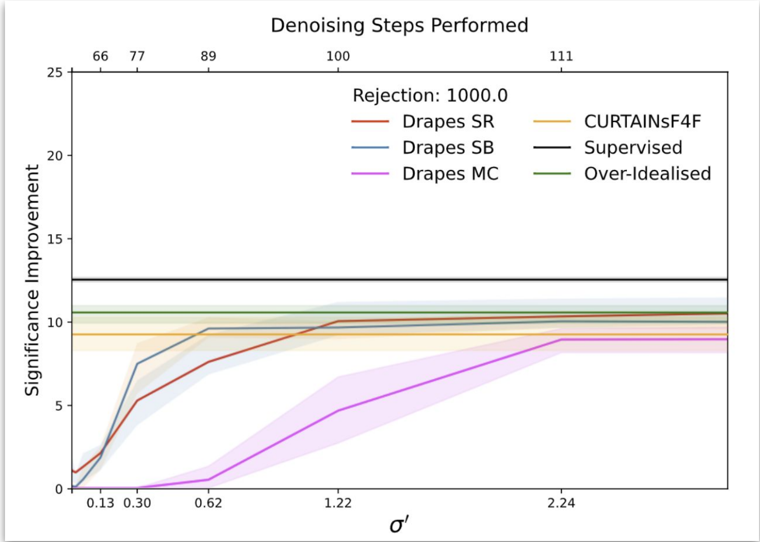


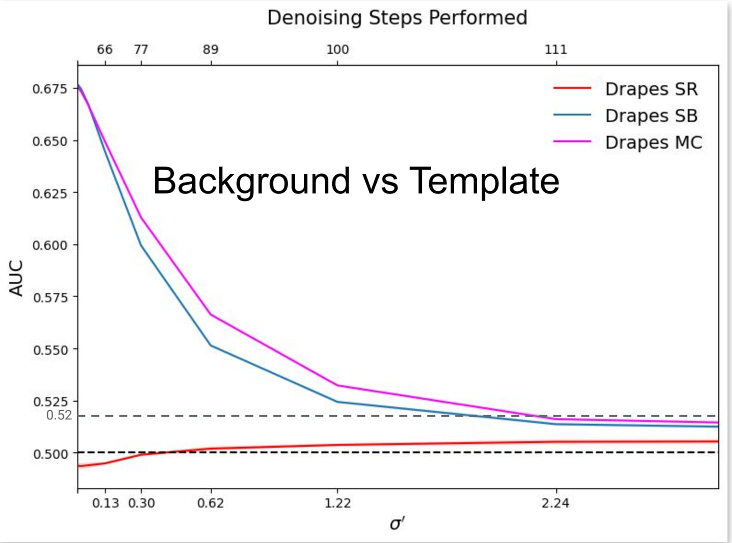
t=1

How much noise is enough?

Also see Stephen Jiggins's talk Friday

Partial Diffusion on High Level





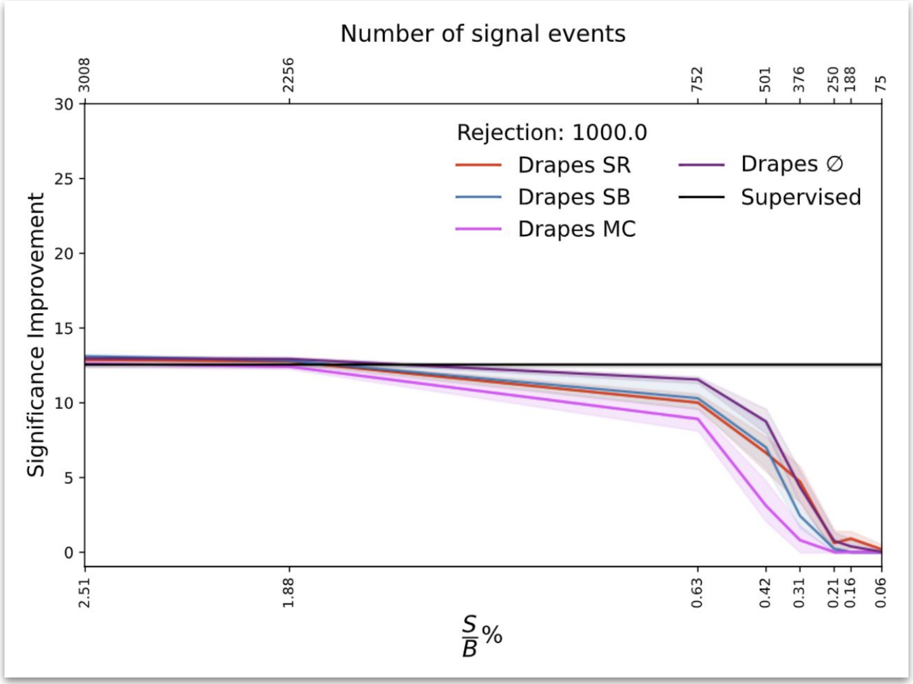
- Method works but SIC performance saturates rather than peaks
- Background vs Template separation ~ 0.515, **even lower for Drapes SR**

Not using variance preserving diffusion so use σ' instead of t; σ_{max} = 80

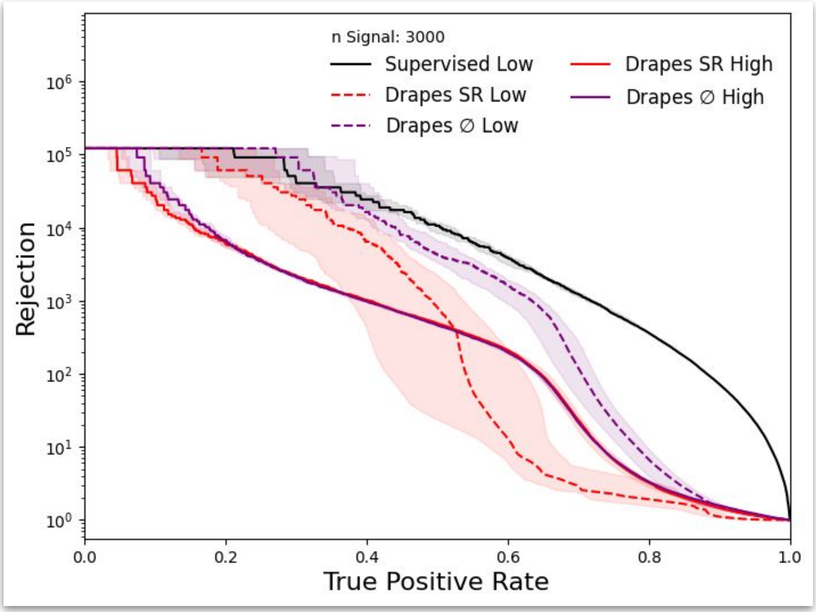
Partial Diffusion on High Level features

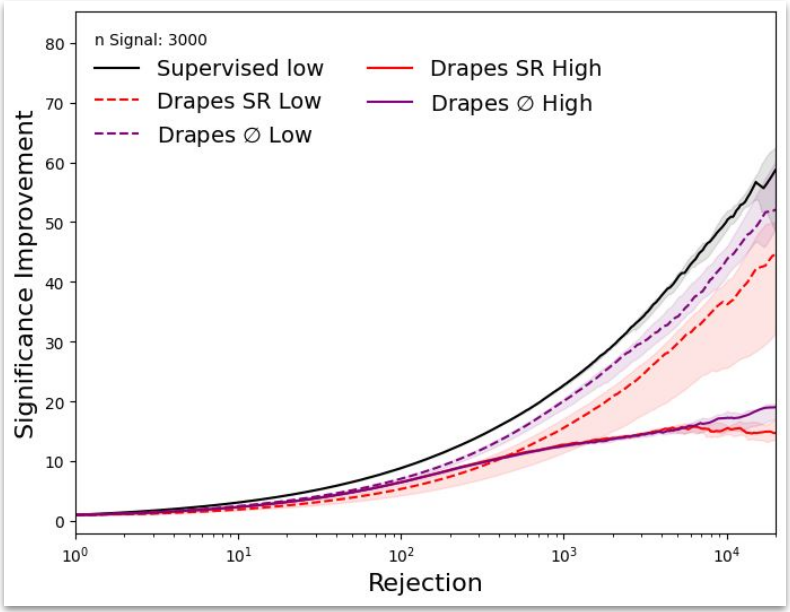
Look at behaviour for range of signal injections with σ' = 2.24

- Background templates in good agreement with data
- Partial methods don't perform as well as full diffusion
    - But reduce overall computation for inference
    - Drapes SR lower AUC in no signal case



Partial Diffusion on Low Level





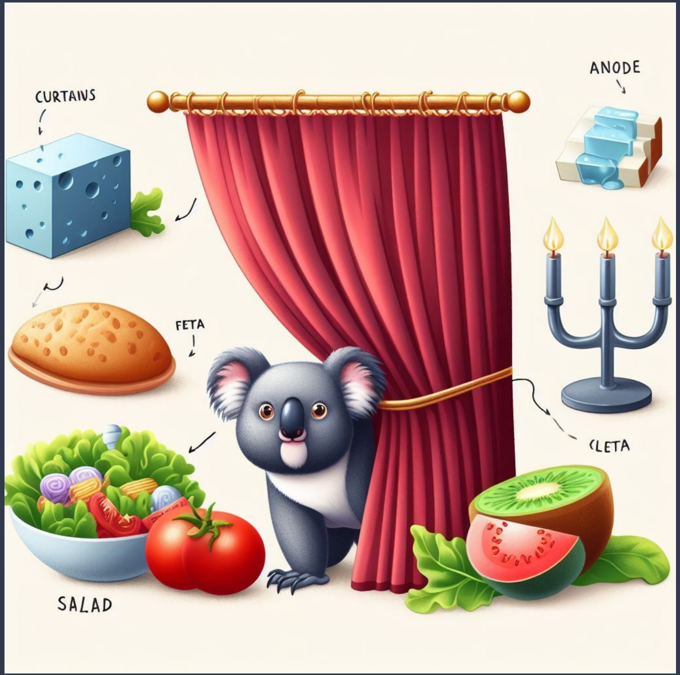
- For **Drapes SR Low** we only modify jet point cloud (not mjj or jet kinematics)
- Performance is not as good as full generation but still outperforms high level!

Conclusions

Stay tuned for arXiv:23XX.XXXX

- Diffusion perfectly viable for template generation
  - Drapes achieves **state of the art** performance on LHCO
- Drapes works really well **using low level** information
  - We do see CWoLa struggling in this setting with low signal
  - We have seen that pre-training may help this
- Drapes with **partial diffusion shows promise**
  - Does not perform as well as full generation
  - Can get more accurate templates in absence of signal
  - Can significantly reduce inference time

Thank You



"Draw an image containing the following: Curtains, drapes, salad, feta, a koala, an anode, and a cathode"

Backup

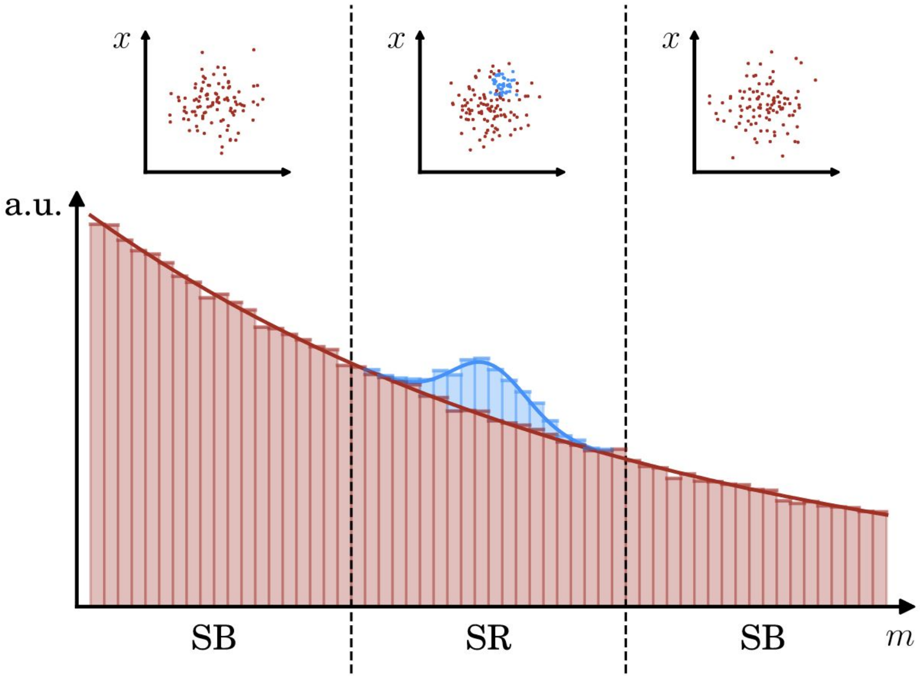
Diffusion Models

- **Forward / noising process**
  - Sample data $p(\mathbf{x}_0)$ → turn to noise

- **Reverse / denoising process**
  - Sample noise $p_T(\mathbf{x}_T)$ → turn into data

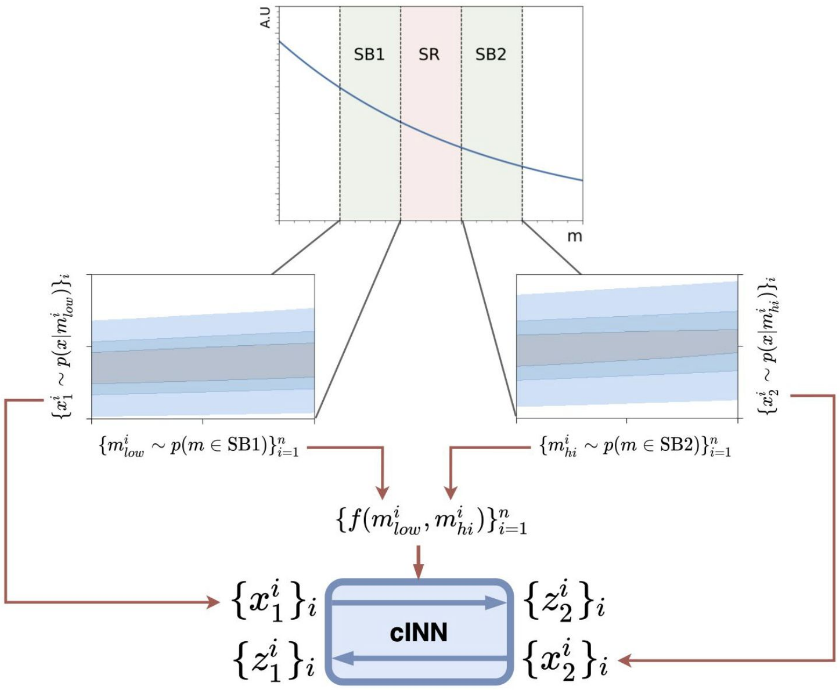
CATHODE

- Use **NORMALISING FLOW**
  - Train on sidebands
  - Condition on mass
  - Use to generate in signal region



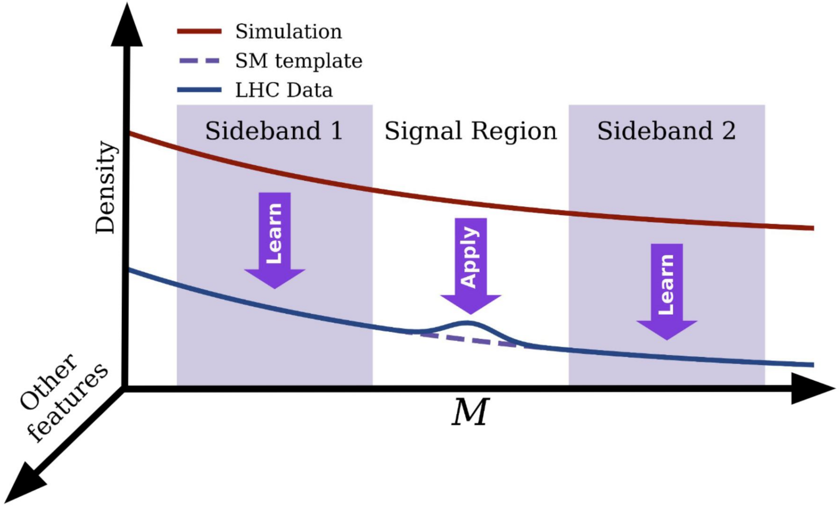
CURTAINS

- Rather than generate from scratch
- Learn how to **modify** data
    - **ie: Take a sample, give it a new mass, and morph**



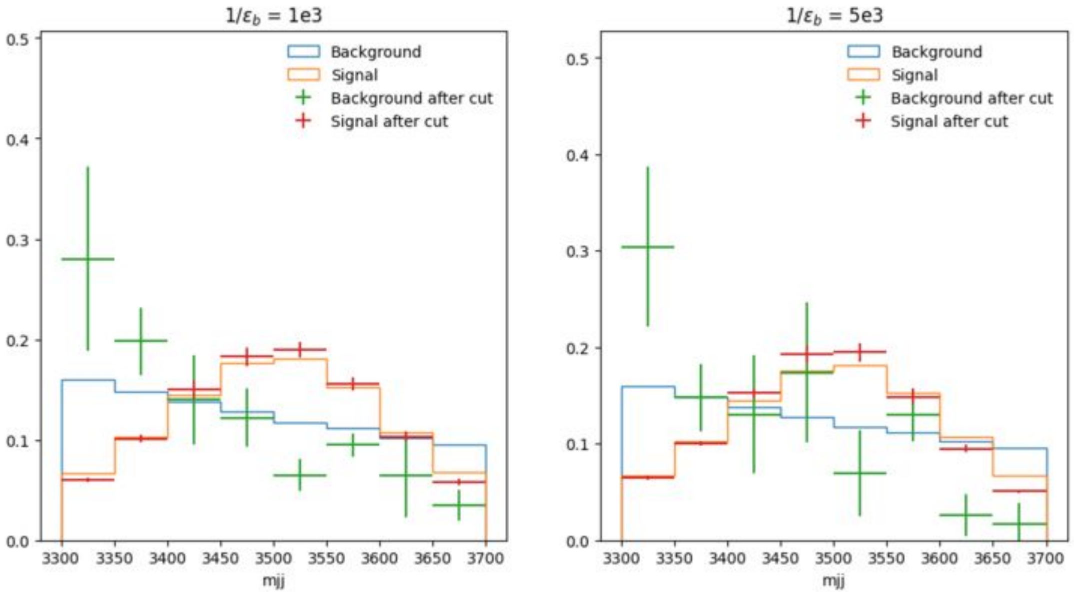
FETA

- Learn to transform **MC** to **DATA**
- Train by transforming sidebands
- Apply in signal region
- Learn how to **modify** data
    - Give it **new origin**

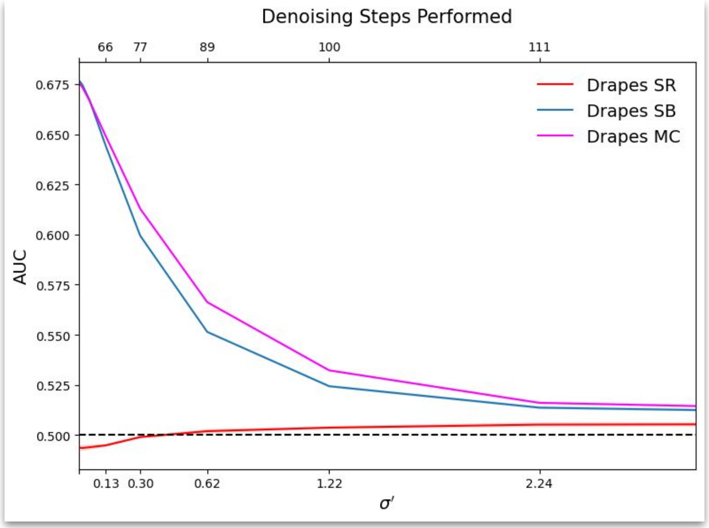


Mass Sculpting

- Verified that the classifier does not appear to significantly sculpt the mass



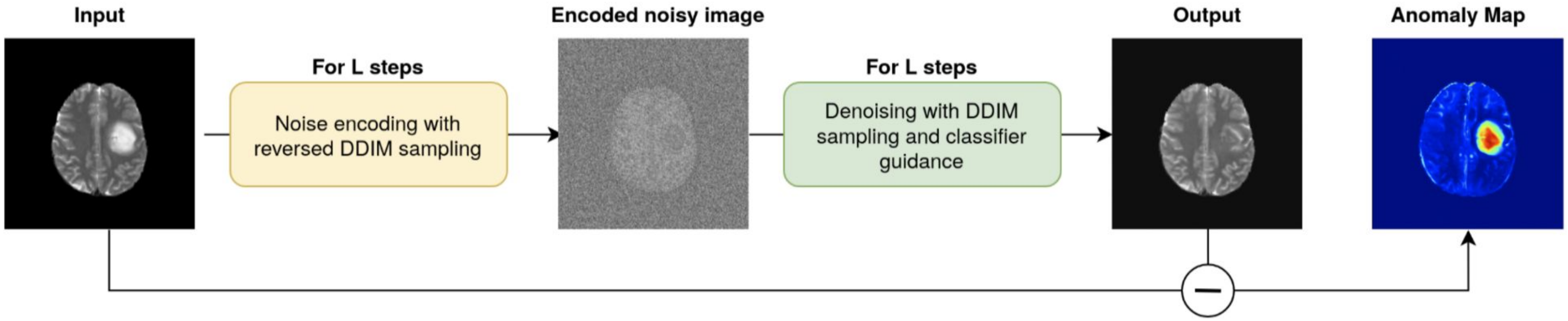
Background Template Seperation



AUC for template vs background as a function of sigma’

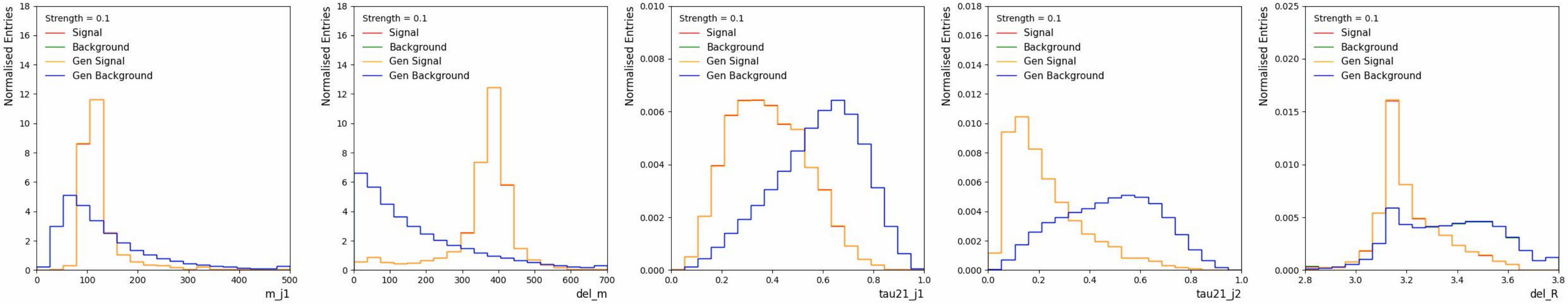
Diffusion Anomaly Detection

- Method has seen success in image applications
- Won't be exactly how we will use it

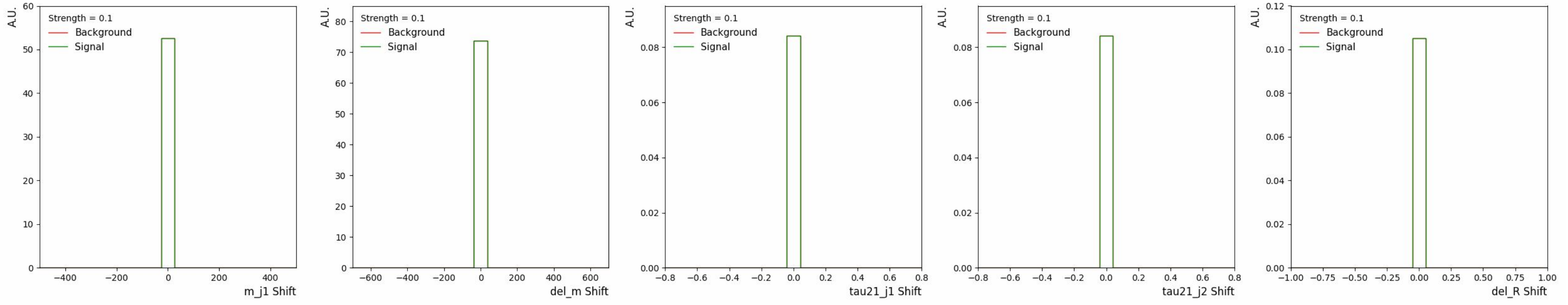


https://arxiv.org/pdf/2203.04306.pdf

Drapes SR – Effect on Distributions



Drapes SR – Effect on Sample



Exclusion Limits