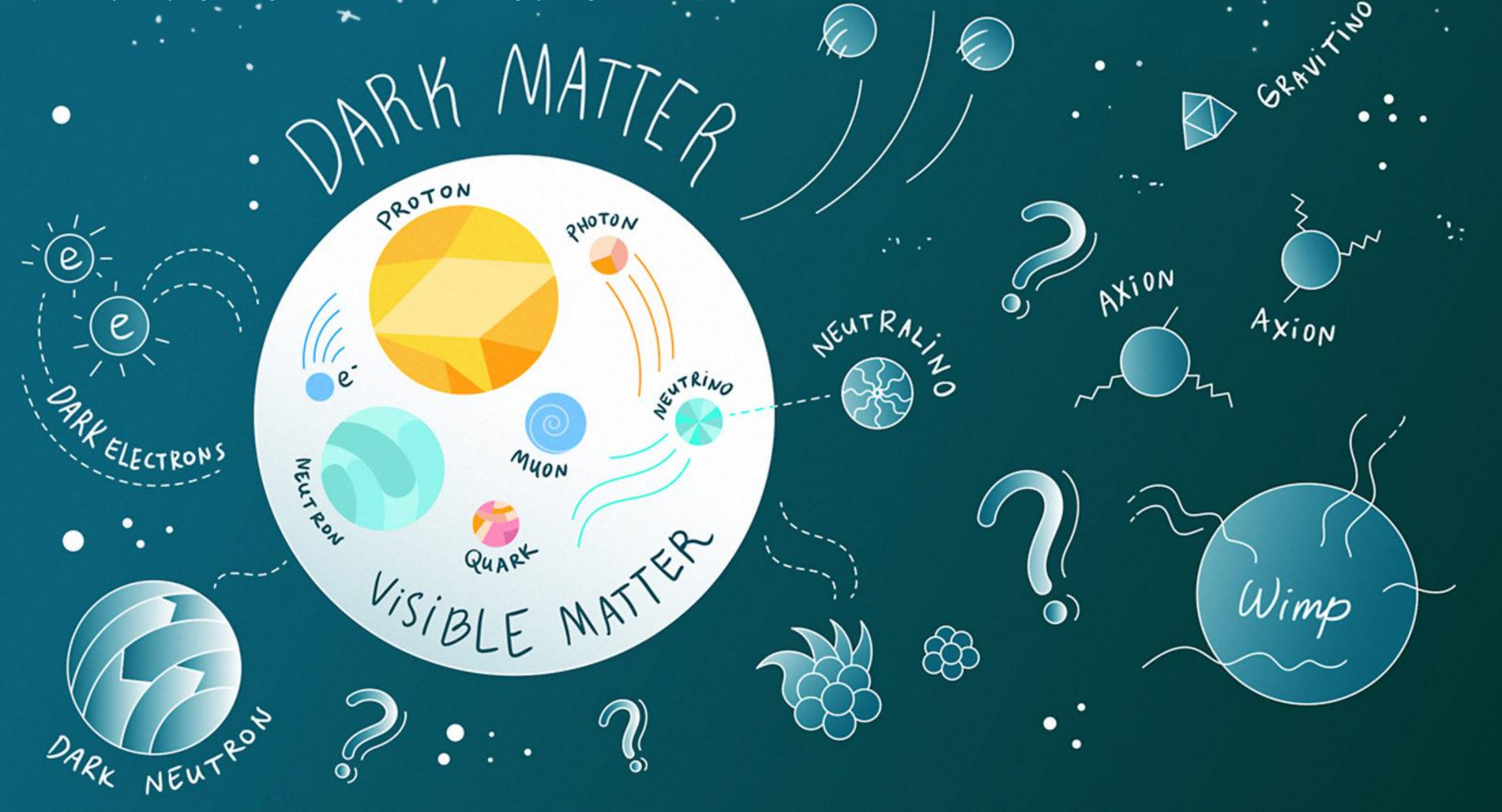
Anomaly Detection: State-of-the-Art

vmikuni@lbl.gov

vinicius-mikuni

Vinicius M. Mikuni

https://www.symmetrymagazine.org/article/december-2013/four-things-you-might-not-know-about-dark-matter



The Challenge

BERKELEY LAB

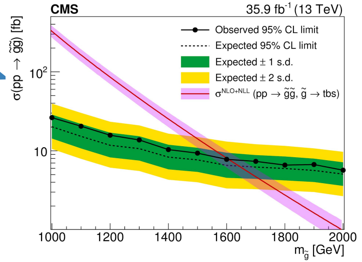
Open Problem: What is dark matter?

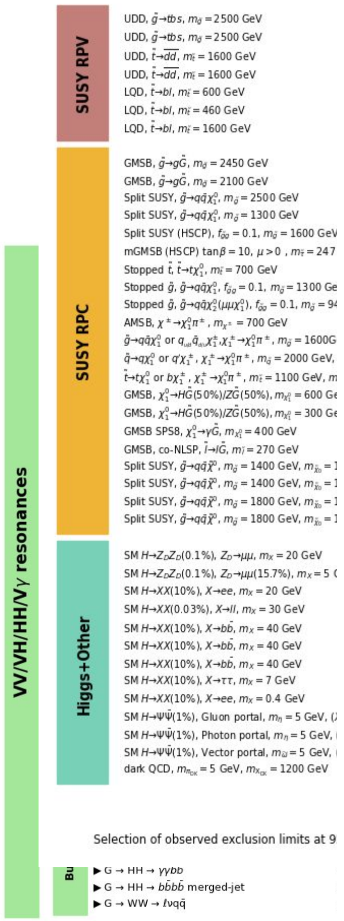
Plausible Theory: SUSY

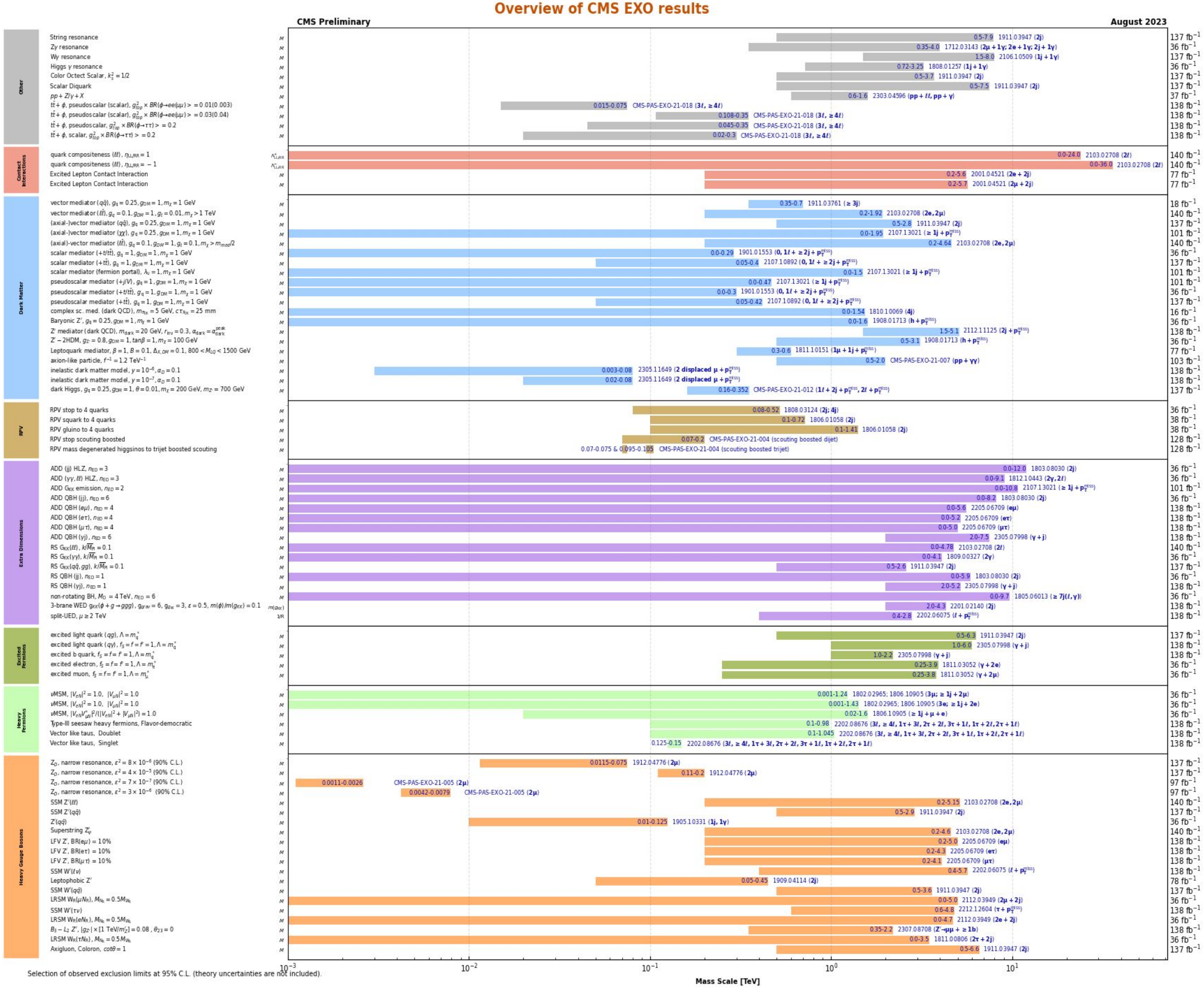
Verification: Confirm the theory using data

Theory was right!

Constrain the new theory




SUSY RPV
UDD, $\tilde{g}\to tbs$, $m_{\tilde{g}} = 2500$ GeV
UDD, $\tilde{g}\to tbs$, $m_{\tilde{g}} = 2500$ GeV
UDD, $\tilde{t}\to\overline{dd}$, $m_{\tilde{t}} = 1600$ GeV
UDD, $\tilde{t}\to\overline{dd}$, $m_{\tilde{t}} = 1600$ GeV
LQD, $\tilde{t}\to bl$, $m_{\tilde{t}} = 600$ GeV
LQD, $\tilde{t}\to bl$, $m_{\tilde{t}} = 460$ GeV
LQD, $\tilde{t}\to bl$, $m_{\tilde{t}} = 1600$ GeV
SUSY RPC
GMSB, $\tilde{g}\to g\tilde{G}$, $m_{\tilde{g}} = 2450$ GeV
GMSB, $\tilde{g}\to g\tilde{G}$, $m_{\tilde{g}} = 2100$ GeV
Split SUSY, $\tilde{g}\to q\bar{q}\tilde{\chi}^0_1$, $m_{\tilde{g}} = 2500$ GeV
Split SUSY, $\tilde{g}\to q\bar{q}\tilde{\chi}^0_1$, $m_{\tilde{g}} = 1300$ GeV
Split SUSY (HSCP), $f_{\tilde{g}g} = 0.1$, $m_{\tilde{g}} = 1600$ GeV
mGMSB (HSCP) $\tan\beta = 10$, $\mu > 0$, $m_{\tilde{\tau}} = 247$
Stopped $\tilde{t}$, $\tilde{t}\to t\tilde{\chi}^0_1$, $m_{\tilde{t}} = 700$ GeV
AMSB, $\chi^{\pm}\to\chi^0_1\pi^{\pm}$, $m_{\chi^{\pm}} = 700$ GeV
GMSB SPS8, $\chi^0_1\to\gamma\tilde{G}$, $m_{\chi^0_1} = 400$ GeV
GMSB, co-NLSP, $\tilde{l}\to l\tilde{G}$, $m_{\tilde{l}} = 270$ GeV
VV/VH/HH/Vγ resonances
Higgs+Other
SM $H\to Z_D Z_D(0.1\%)$, $Z_D\to\mu\mu$, $m_X = 20$ GeV
SM $H\to XX(10\%)$, $X\to ee$, $m_X = 20$ GeV
SM $H\to XX(0.03\%)$, $X\to ll$, $m_X = 30$ GeV
SM $H\to XX(10\%)$, $X\to b\bar{b}$, $m_X = 40$ GeV
SM $H\to XX(10\%)$, $X\to\tau\tau$, $m_X = 7$ GeV
SM $H\to XX(10\%)$, $X\to ee$, $m_X = 0.4$ GeV
dark QCD, $m_{\pi_{DK}} = 5$ GeV, $m_{X_{DK}} = 1200$ GeV
Selection of observed exclusion limits at 9
▶ G → HH → $\gamma\gamma b\bar{b}$
▶ G → HH → $b\bar{b}b\bar{b}$ merged-jet
▶ G → WW → $\ell\nu q\bar{q}$

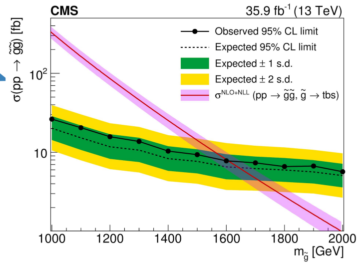

Overview of CMS EXO results
CMS Preliminary
August 2023
Other
Contact Interactions
Dark Matter
RPV
Extra Dimensions
Excited Fermions
Heavy Fermions
Heavy Gauge Bosons
Selection of observed exclusion limits at 95% C.L. (theory uncertainties are not included).
Mass Scale [TeV]

The Challenge

Open Problem: What is dark matter? → **Plausible Theory**: SUSY → **Verification**: Confirm the theory using data

Theory was right!

Constrain the new theory

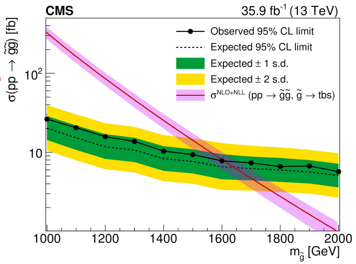


The Challenge

Look for anomalies

Interpretation

Constrain many theories



The Challenge

https://iml-wg.github.io/HEPML-LivingReview/#anomaly-detection

Anomaly detection. ¶

Anomaly detection.

- Learning New Physics from a Machine [DOI]
- Anomaly Detection for Resonant New Physics with Machine Learning [DOI]
- Extending the search for new resonances with machine learning [DOI]
- Learning Multivariate New Physics [DOI]
- Searching for New Physics with Deep Autoencoders [DOI]
- QCD or What? [DOI]
- A robust anomaly finder based on autoencoder
- Variational Autoencoders for New Physics Mining at the Large Hadron Collider [DOI]
- Adversarially-trained autoencoders for robust unsupervised new physics searches [DOI]
- Novelty Detection Meets Collider Physics [DOI]
- Guiding New Physics Searches with Unsupervised Learning [DOI]
- Does SUSY have friends? A new approach for LHC event analysis [DOI]
- Nonparametric semisupervised classification for signal detection in high energy physics
- Uncovering latent jet substructure [DOI]
- Simulation Assisted Likelihood-free Anomaly Detection [DOI]
- Anomaly Detection with Density Estimation [DOI]
- A generic anti-QCD jet tagger [DOI]
- Transferability of Deep Learning Models in Searches for New Physics at Colliders [DOI]
- Use of a Generalized Energy Mover's Distance in the Search for Rare Phenomena at Colliders [DOI]
- Adversarially Learned Anomaly Detection on CMS Open Data: re-discovering the top quark [DOI]
- Dijet resonance search with weak supervision using 13 TeV pp collisions in the ATLAS detector [DOI]
- Learning the latent structure of collider events [DOI]
- Finding New Physics without learning about it: Anomaly Detection as a tool for Searches at Colliders [DOI]
- Tag N' Train: A Technique to Train Improved Classifiers on Unlabeled Data [DOI]
- Variational Autoencoders for Anomalous Jet Tagging [DOI]
- Anomaly Awareness [DOI]
- Unsupervised Outlier Detection in Heavy-Ion Collisions [DOI]
- Decoding Dark Matter Substructure without Supervision
- Mass Unspecific Supervised Tagging (MUST) for boosted jets [DOI]
- Simulation-Assisted Decorrelation for Resonant Anomaly Detection [DOI]
- Anomaly Detection With Conditional Variational Autoencoders
- Unsupervised clustering for collider physics [DOI]
- Combining outlier analysis algorithms to identify new physics at the LHC [DOI]
- Quasi Anomalous Knowledge: Searching for new physics with embedded knowledge [DOI]
- Uncovering hidden patterns in collider events with Bayesian probabilistic models [DOI]
- Unsupervised in-distribution anomaly detection of new physics through conditional density estimation
- The LHC Olympics 2020: A Community Challenge for Anomaly Detection in High Energy Physics [DOI]
- Model-Independent Detection of New Physics Signals Using Interpretable Semi-Supervised Classifier Tests
- Topological Obstructions to Autoencoding [DOI]
- Unsupervised Event Classification with Graphs on Classical and Photonic Quantum Computers [DOI]
- Bump Hunting in Latent Space [DOI]
- Comparing Weak- and Unsupervised Methods for Resonant Anomaly Detection [DOI]
- Better Latent Spaces for Better Autoencoders [DOI]
- Autoencoders for unsupervised anomaly detection in high energy physics [DOI]
- Via Machinae: Searching for Stellar Streams using Unsupervised Machine Learning [DOI]
- Anomaly detection with Convolutional Graph Neural Networks [DOI]
- Anomalous Jet Identification via Sequence Modeling [DOI]
- The Dark Machines Anomaly Score Challenge: Benchmark Data and Model Independent Event Classification for the Large Hadron Collider [DOI]

- RanBox: Anomaly Detection in the Copula Space [DOI]
- Rare and Different: Anomaly Scores from a combination of likelihood and out-of-distribution models to detect new physics at the LHC [DOI]
- LHC physics dataset for unsupervised New Physics detection at 40 MHz [DOI]
- New Methods and Datasets for Group Anomaly Detection From Fundamental Physics
- The Data-Directed Paradigm for BSM searches [DOI]
- Autoencoders on FPGAs for real-time, unsupervised new physics detection at 40 MHz at the Large Hadron Collider [DOI]
- Classifying Anomalies THrough Outer Density Estimation (CATHODE) [DOI]
- Deep Set Auto Encoders for Anomaly Detection in Particle Physics [DOI]
- Challenges for Unsupervised Anomaly Detection in Particle Physics [DOI]
- Improving Variational Autoencoders for New Physics Detection at the LHC with Normalizing Flows [DOI]
- Signal-agnostic dark matter searches in direct detection data with machine learning [DOI]
- Anomaly detection from mass unspecific jet tagging [DOI]
- A method to challenge symmetries in data with self-supervised learning [DOI]
- Stressed GANs snag desserts, a.k.a Spotting Symmetry Violation with Symmetric Functions
- Online-compatible Unsupervised Non-resonant Anomaly Detection [DOI]
- Event-based anomaly detection for new physics searches at the LHC using machine learning [DOI]
- Learning New Physics from an Imperfect Machine [DOI]
- Autoencoders for Semivisible Jet Detection [DOI]
- Anomaly detection in high-energy physics using a quantum autoencoder [DOI]
- Creating Simple, Interpretable Anomaly Detectors for New Physics in Jet Substructure [DOI]
- Taming modeling uncertainties with Mass Unspecific Supervised Tagging [DOI]
- What's Anomalous in LHC Jets? [DOI]
- Quantum Anomaly Detection for Collider Physics [DOI]
- Detecting new physics as novelty \textemdash{} Complementarity matters [DOI]
- Self-supervised Anomaly Detection for New Physics [DOI]
- Data-directed search for new physics based on symmetries of the SM [DOI]
- CURTAINs for your Sliding Window: Constructing Unobserved Regions by Transforming Adjacent Intervals [DOI]
- Learning new physics efficiently with nonparametric methods [DOI]
- "Flux+Mutability": A Conditional Generative Approach to One-Class Classification and Anomaly Detection [DOI]
- Boosting mono-jet searches with model-agnostic machine learning [DOI]
- Event Generation and Density Estimation with Surjective Normalizing Flows [DOI]
- A Normalized Autoencoder for LHC Triggers [DOI]
- Mixture-of-theories Training: Can We Find New Physics and Anomalies Better by Mixing Physical Theories? [DOI]
- Neural Embedding: Learning the Embedding of the Manifold of Physics Data [DOI]
- Null Hypothesis Test for Anomaly Detection [DOI]
- Resonant anomaly detection without background sculpting [DOI]

- Anomaly Detection under Coordinate Transformations [DOI]
- Quantum-probabilistic Hamiltonian learning for generative modelling \& anomaly detection [DOI]
- Efficiently Moving Instead of Reweighting Collider Events with Machine Learning
- Unravelling physics beyond the standard model with classical and quantum anomaly detection [DOI]
- Nanosecond anomaly detection with decision trees for high energy physics and real-time application to exotic Higgs decays [DOI]
- The Mass-ive Issue: Anomaly Detection in Jet Physics
- CURTAINs Flows For Flows: Constructing Unobserved Regions with Maximum Likelihood Estimation
- High-dimensional and Permutation Invariant Anomaly Detection [DOI]
- The Interplay of Machine Learning--based Resonant Anomaly Detection Methods [DOI]
- GAN-AE : An anomaly detection algorithm for New Physics search in LHC data [DOI]
- Anomaly detection search for new resonances decaying into a Higgs boson and a generic new particle X in hadronic final states using $\sqrt(s) [DOI]
- Boosting sensitivity to new physics with unsupervised anomaly detection in dijet resonance search [DOI]
- Autoencoder-based Anomaly Detection System for Online Data Quality Monitoring of the CMS Electromagnetic Calorimeter
- Combining Resonant and Tail-based Anomaly Detection
- Back To The Roots: Tree-Based Algorithms for Weakly Supervised Anomaly Detection [DOI]
- Full Phase Space Resonant Anomaly Detection [DOI]
- Anomaly Detection in Presence of Irrelevant Features [DOI]
- Triggerless data acquisition pipeline for Machine Learning based statistical anomaly detection [DOI]
- Non-resonant Anomaly Detection with Background Extrapolation [DOI]
- Searching for gluon quartic gauge couplings at muon colliders using the auto-encoder
- Fast Particle-based Anomaly Detection Algorithm with Variational Autoencoder
- Anomaly Detection in Collider Physics via Factorized Observables
- Testing a Neural Network for Anomaly Detection in the CMS Global Trigger Test Crate during Run 3 [DOI]
- Improving new physics searches with diffusion models for event observables and jet constituents [DOI]
- Anomaly detection with flow-based fast calorimeter simulators
- Incorporating Physical Priors into Weakly-Supervised Anomaly Detection
- Accelerating Resonance Searches via Signature-Oriented Pre-training

100+ anomaly detection papers so far

Common approaches

BERKELEY LAB

Signal is rare but so is the background

- NP at “**tails**” of distributions
- **Challenges**: **“rarity” is not universal** and the choice of features determine the sensitivity

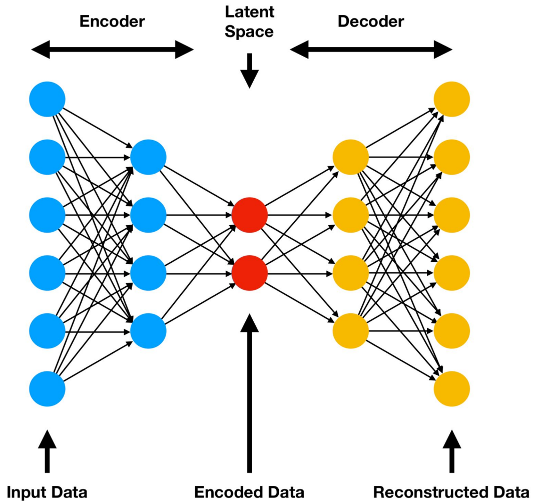
Signal is an over density:

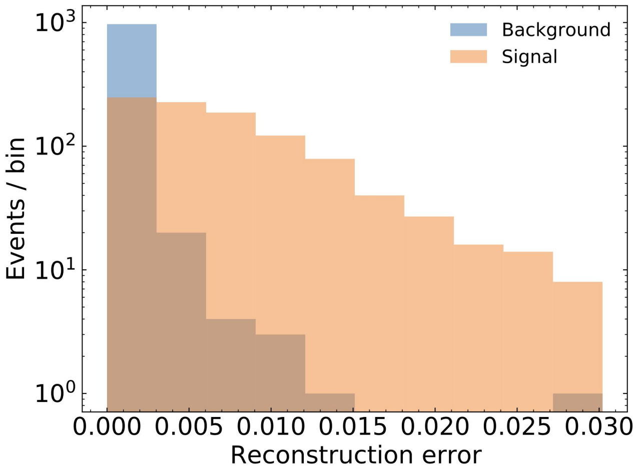
- New physics is rare, but at least one feature has a region where **$p_s(x)/p_b(x)$ is high**
- **Challenges**: Requires an estimate of $p_b(x)$ and prior knowledge of the resonant feature

Non-resonant anomaly detection

Signal is rare but so is the background

- NP at "**tails**" of distributions
- **Challenges**: "**rarity" is not universal** and the choice of features determine the sensitivity

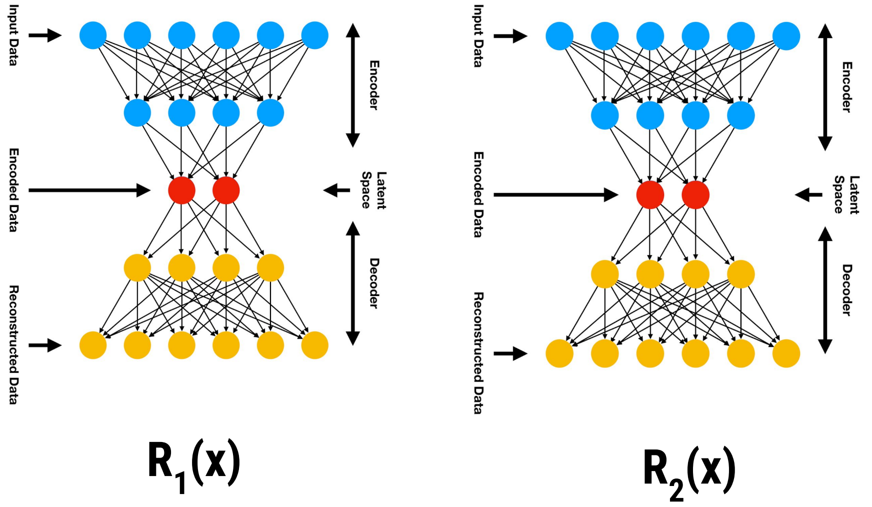




Double Autoencoder

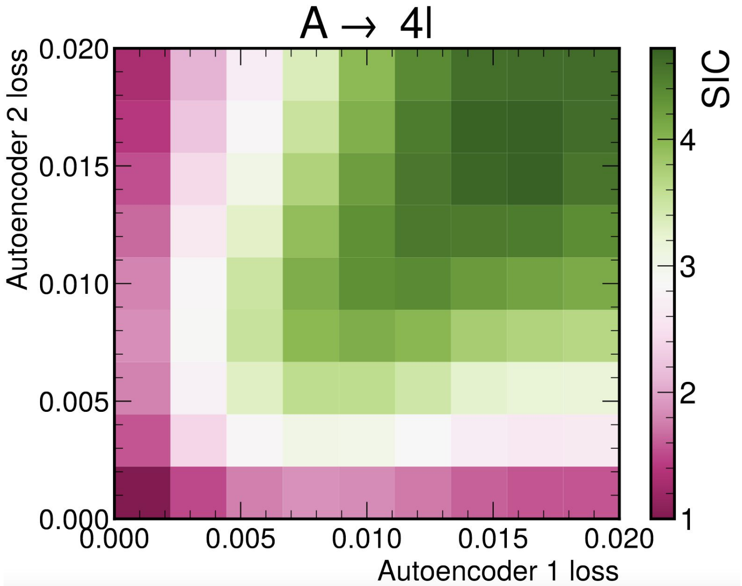
- Train **multiple autoencoders** and enforce decorrelation for background events

$$L[f_1, f_2, g_1, g_2] = \sum_i R_1(x_i)^2 + \sum_i R_2(x_i)^2 + \lambda\,\mathrm{DisCo}^2[R_1(X), R_2(X)]$$



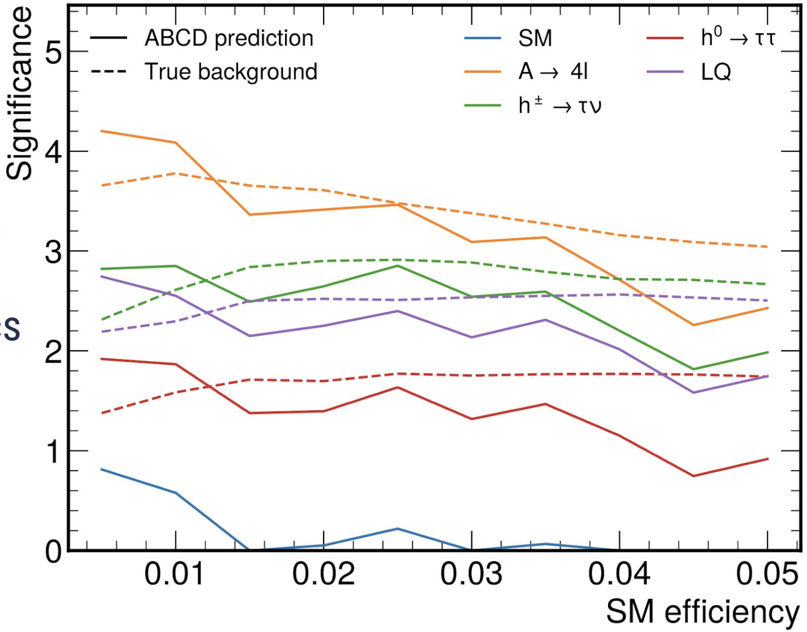
V. Mikuni, B. Nachman, and D. Shih Phys. Rev. D 105, 055006

Anomaly detection performance



No anomalies

Other colors: datasets with **0.1%** anomalies and **99.9%** standard physics processes



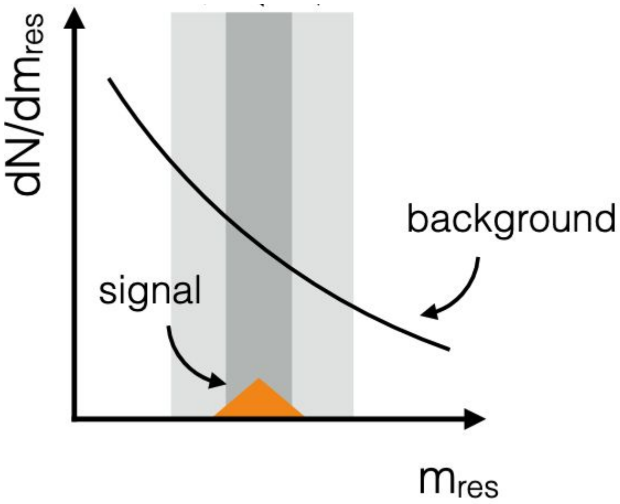
- Use the ABCD method to determine the background

V. Mikuni, B. Nachman, and D. Shih Phys. Rev. D 105, 055006

Resonant anomaly detection

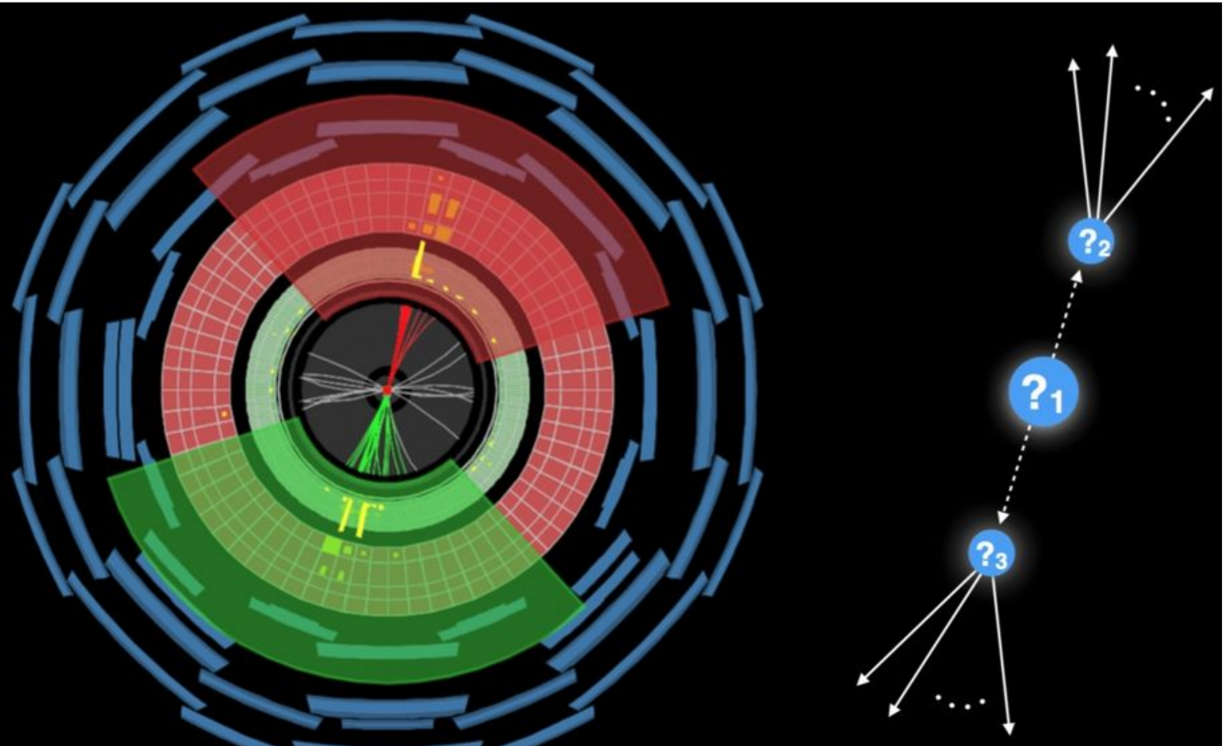
Signal is an over density:

- New physics is rare, but at least one feature has a region where **$p_s(x)/p_b(x)$ is high**
- **Challenges**: Requires an estimate of $p_b(x)$ and prior knowledge of the resonant feature



- **Bump hunts** often only use the resonant feature to achieve sensitivity
- ML enables the use of **multiple features** while maintaining the benefits of a resonant signal to perform background estimation

LHCO dataset

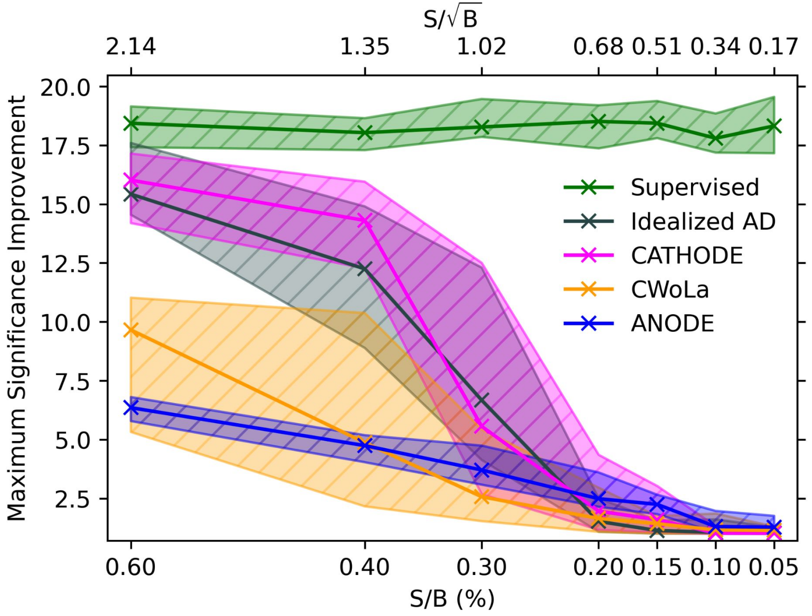


Detector cross-section
image credit: ATLAS

- Common benchmark for resonant AD: **LHCO R&D dataset**
- Resonant **dijet** final state: A->B(qq)C(qq) with m_A, m_B , m_C = 3.5, 0.5, 0.1 TeV

CATHODE

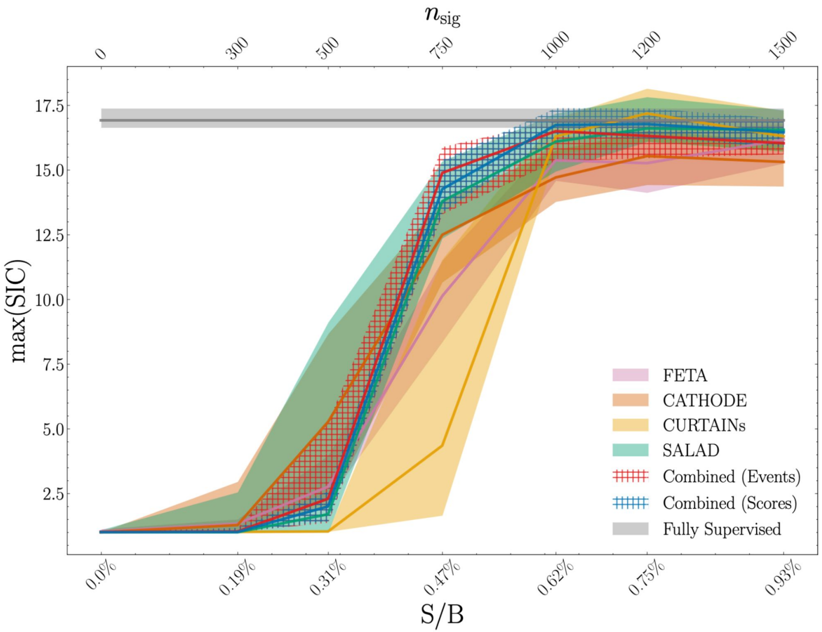
- Learn a **4-dimensional** background using the sidebands
  - N-subjettiness
  - Jet masses
- **Interpolate** the background prediction in the signal region
- Train a classifier to distinguish the interpolated background from the data
- **SIC** = Significance Improvement Curve (TPR/sqrt(FPR) vs TPR) “By how much can I improve the significance of a particular signal given an initial significance.”



A. Hallin, J. Isaacson, G. Kasieczka, C. Krause, B. Nachman, T. Quadfasel, M. Schlaffer, D. Shih, and M. Sommerhalder, Phys. Rev. D 106, 055006

Combining multiple methods

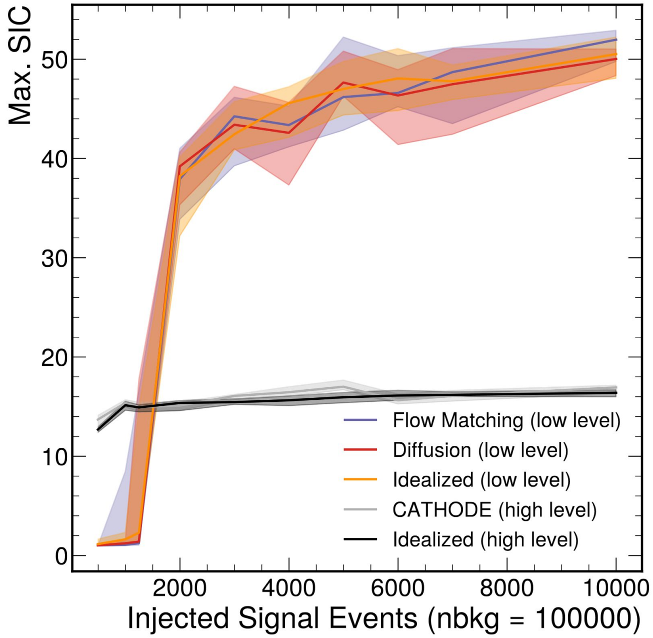
- Many strategies available to determine the background distribution
- Even though the performance of different methods is similar, they are **not** completely correlated
- Compare and investigate different combination strategies: smaller uncertainty bands!



Golling, T., Kasieczka, G., Krause, C. et al. Eur. Phys. J. C 84, 241 (2024).

Going above and beyond

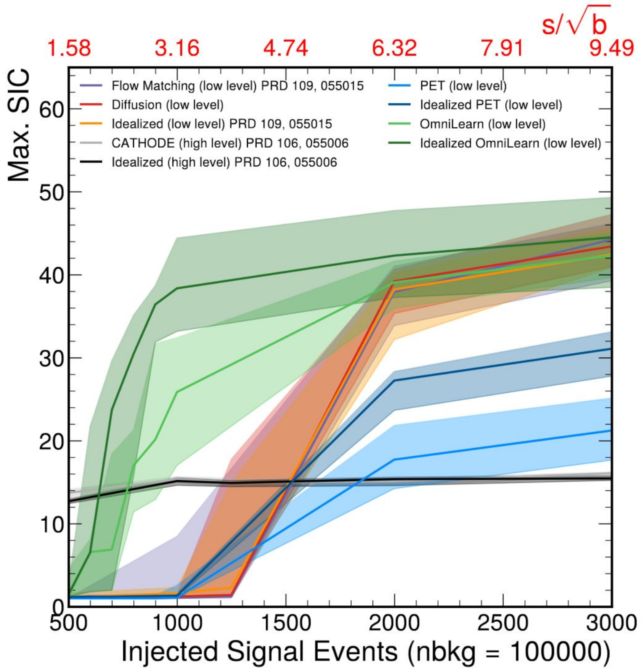
- Current methods are limited to the 4 features
- Strong performance for new physics scenarios where **jet substructure** is important
- What if we estimate the complete dijet system?
    - 4 features vs up to 2*279*3 = **1674 features**
- Method still works and outperforms the method with 4 features for **S/B>3%**



E. Buhmann, C. Ewen, G. Kasieczka, **V. Mikuni**, B. Nachman, and D. Shih, Phys. Rev. D 109, 055015

Going above and beyond

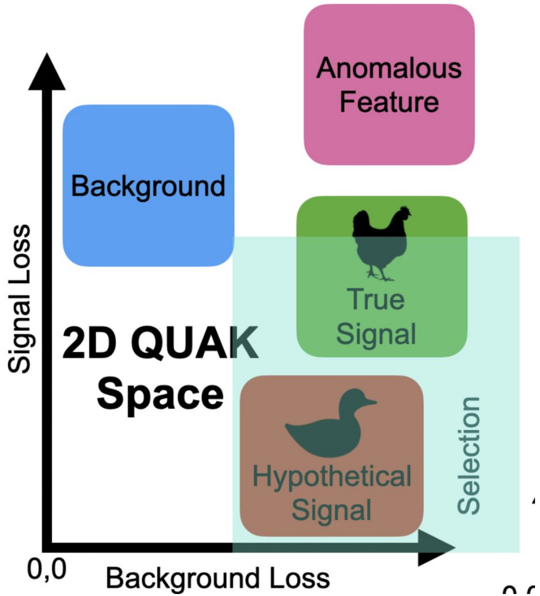
- In this very high dimensional space, performance is limited by the data
- Data cannot be easily increased: requires longer data collection periods
- What if we had a model that understands jets and only asked to adapt to this particular phase space?
- Using **OmniLearn**, a foundational model for jet physics, we are able to be sensitive to the new physics signal with as little as **S/B = 0.7% ~ 2σ**

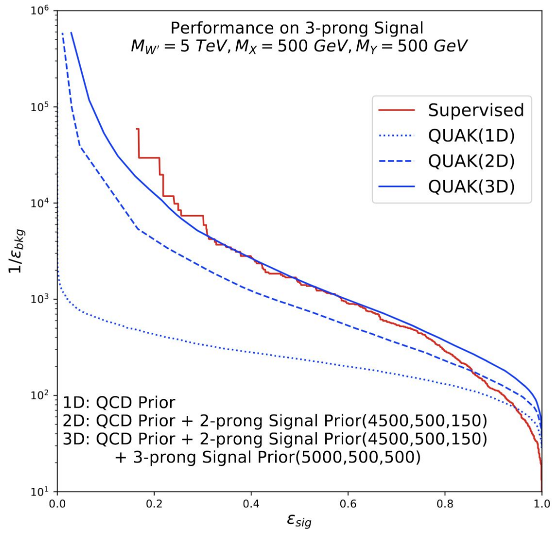


V. **Mikuni,** B. Nachman, arXiv:2404.16091

Other assumptions

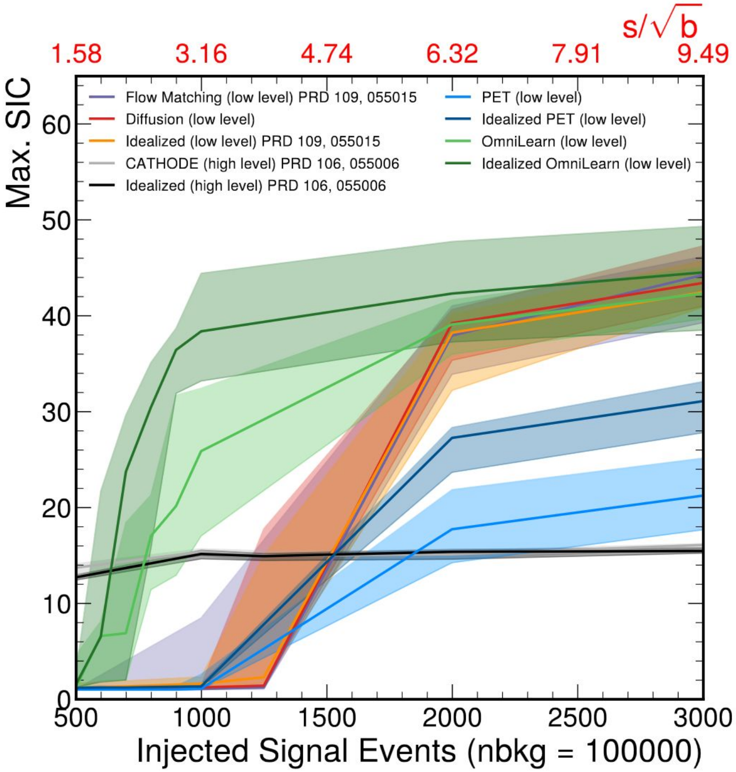
- Additional assumptions beyond the 2 main classes have also been proposed
- Example: **QUAK**
    - Train a model using **multiple** possible NP scenarios to define ”signal-like” and “background-like” regions of the phase space



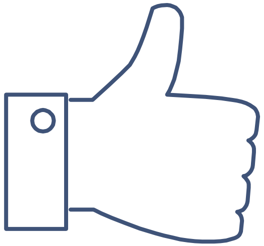


Park, S.E., Rankin, D., Udrescu, SM. et al. J. High Energ. Phys. 2021, 30 (2021).

Conclusions



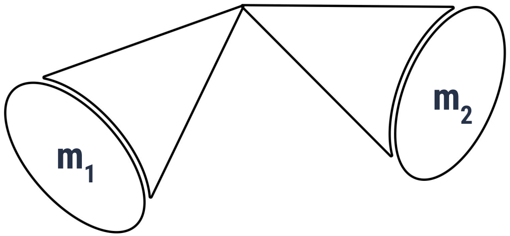
- **Anomaly detection** is an alternative and complementary strategy to search for new physics processes
- **Not magic**: Different anomaly detection methods rely on a few assumptions and is important to be aware of their limitations
- Most methods are **data-driven**: the bigger the dataset the better
- **Foundational models** might be able to bridge the sensitivity gap and allow sensitivity even in low data regimes

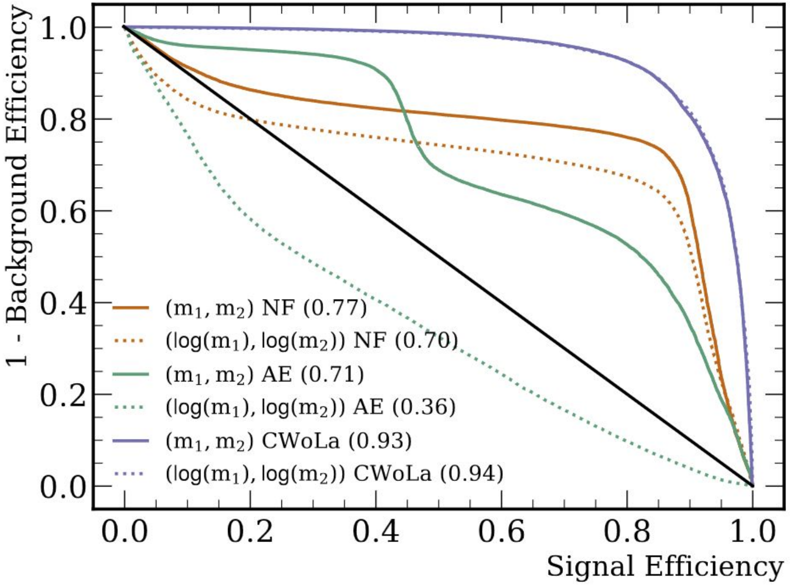


THANKS!

Any questions?

Feature Dependence

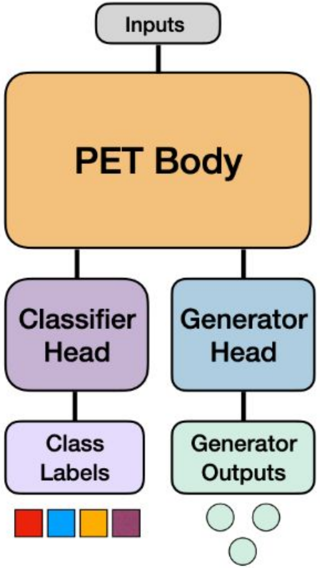




- Choice of **representation of inputs** also affects the performance!
- Differences in performance for autoencoders when using **m_1, m_2** as inputs or **$\log(m_1)$, $\log(m_2)$**

Kasieczka, G., Mastandrea, R., **Mikuni, V.**, Nachman, B., Pettee, M., & Shih, D. (2023). *Physical Review D*, *107*(1), 015009.

OmniLearn



- Train a transformer model to classify and generate jets in the JetClass dataset
  - 100M jets with 10 different jet classes
- Use this pre-trained model as the starting point for multiple tasks
  - **9 additional datasets** tested, including ATLAS top tagging, CMS Open data, and electron-proton collisions
  - **11 applications** including standard classification, generation, anomaly detection, and unfolding
- OmniLearn improves upon **all** tasks investigated with faster convergence

V. Mikuni, B. Nachman, arXiv:2404.16091

Diffusion Generative Models

Forward SDE (data → noise)

$$\mathbf{x}(0) \longrightarrow \mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\mathrm{d}t + g(t)\mathrm{d}\mathbf{w} \longrightarrow \mathbf{x}(T)$$

score function

$$\mathbf{x}(0) \longleftarrow \mathrm{d}\mathbf{x} = \left[\mathbf{f}(\mathbf{x}, t) - g^2(t)\boxed{\nabla_{\mathbf{x}} \log p_t(\mathbf{x})}\right]\mathrm{d}t + g(t)\mathrm{d}\bar{\mathbf{w}} \longleftarrow \mathbf{x}(T)$$

Reverse SDE (noise → data)

Source: https://yang-song.net/blog/2021/score/