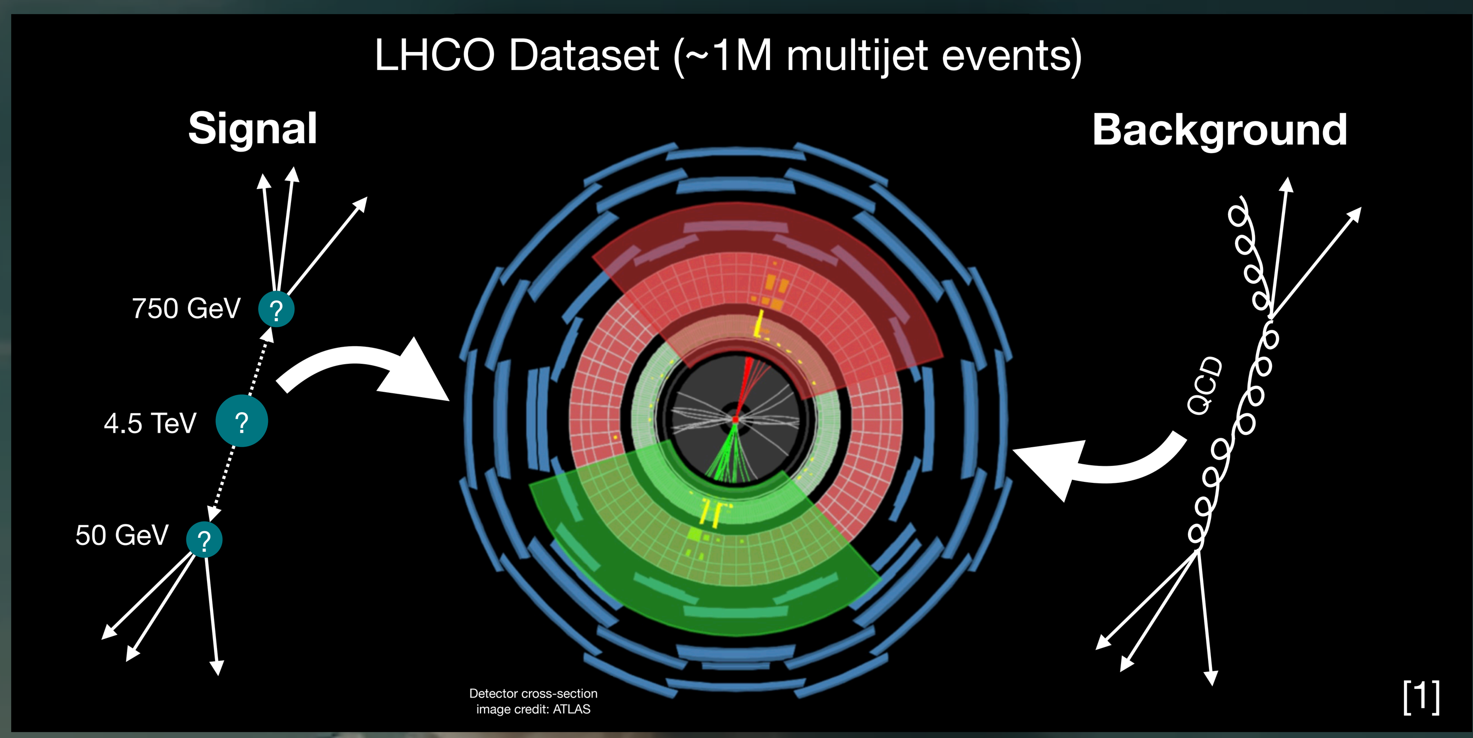
HyLAnD

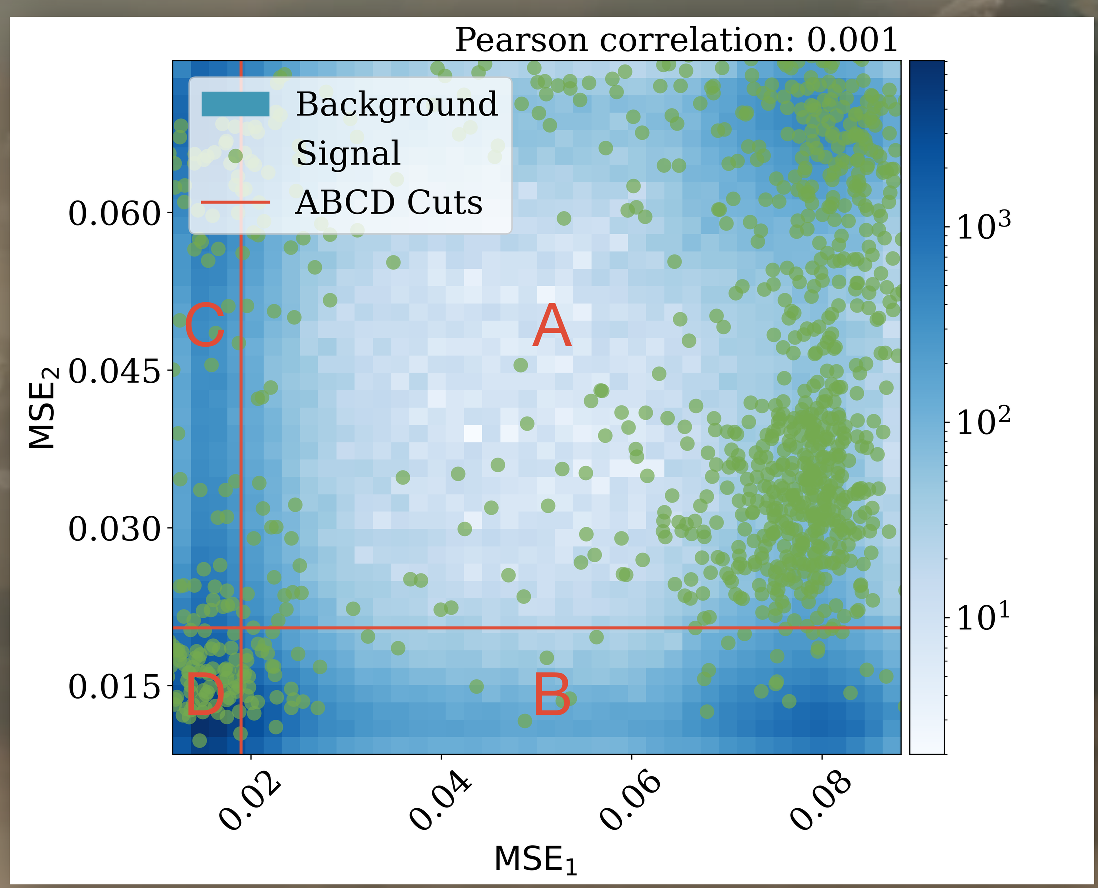
Hybrid Learning for Anomaly Detection

Vinicius Mikuni, Benjamin Nachman, **Dennis Noll**



Motivation

- Want to find new physics in LHC collider data
- Search for anomalies in agnostic way
- Target resonant and non-resonant anomalies
- Study performed on LHCO 2020 dataset [1]



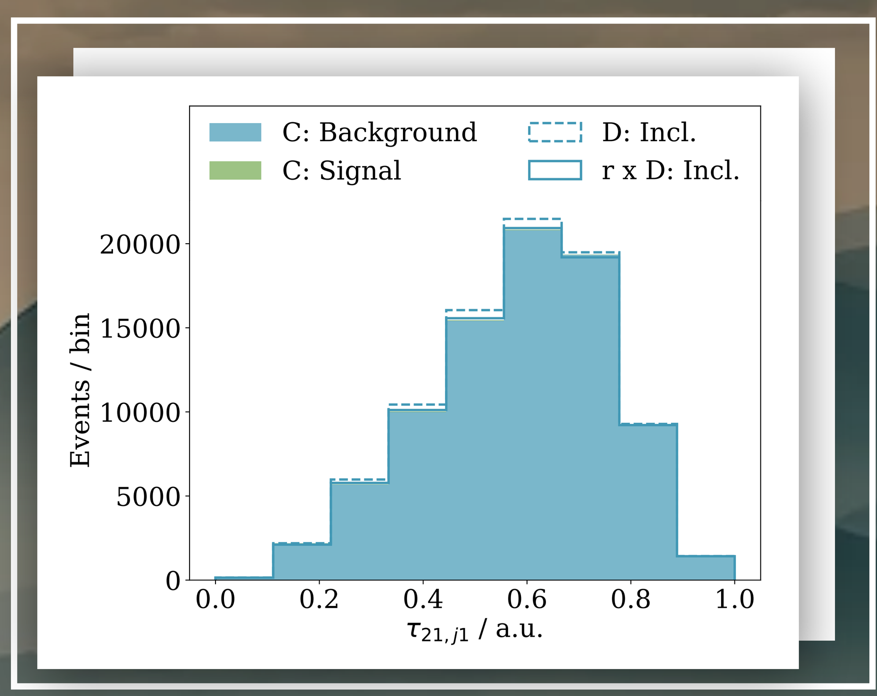
1. Define Signal Region

- Training two decorrelated autoencoders:

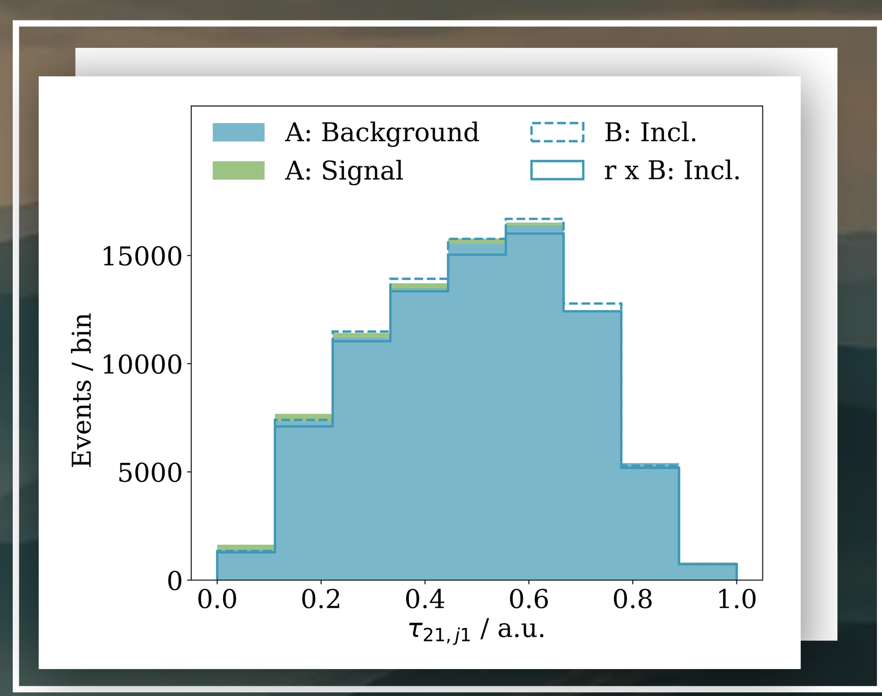
$$\mathscr{L} \propto \mathrm{MSE}_1 + \mathrm{MSE}_2 + \mathrm{DisCo}(\mathrm{MSE}_1, \mathrm{MSE}_2)$$

  - MSE (Resonstruction loss): Finds low densities
  - DisCo (Distance Correlation [2]): Decorrelates AEs
  - Inputs: Kinematic features (m_{jj}, m_{j1}, m_{j2}…)
- **Get:** SR (25% of events with highest MSEs)

CR

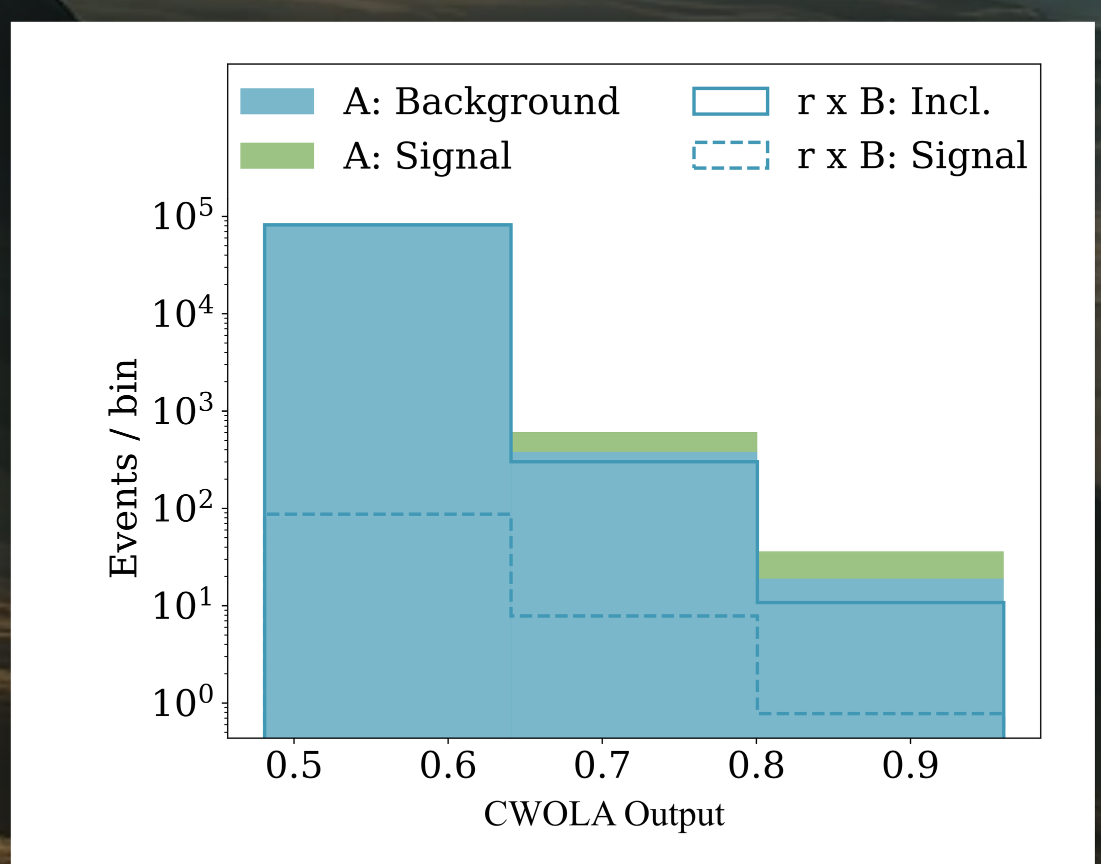


SR



2. Estimate Background

- Use ABCD Method [3]: C / D = A / B
- Differential transfer (R) via classification:
  - Training in Control Region (CR): r = C / D
  - Application in Signal Region (SR): A = r x B
  - Inputs: Substructure features ($\tau_{21,j1}$, $\tau_{32,j1}$, …)
- **Get**: Differential background-only estimate in SR



3. Measure Signal

- Semi-supervised CWOLA [4] in SR:
  - Classification of Data vs Background-estimate
  - Provides optimized discriminant
  - Inputs: Substructure features ($\tau_{21,j1}$, $\tau_{32,j1}$, …)
- **Get** (injected 3σ signal):
  - Highly sensitive signal classifier
  - WIP: further mitigate false positives

Conclusion

- New anomaly detection method for resonant and non-resonant signals
- Hybrid Approach:
  - Unsupervised learning to define signal region and differential bkg. estimate
  - Semi-supervised learning for optimized signal classification
- Can greatly improve signal significance

[1]: G. Kasieczka et al., *The LHC Olympics 2020: A Community Challenge for Anomaly Detection in High Energy Physics*, arXiv:2101.08320
[2]: G. Kasieczka, B. Nachman, M. D. Schwartz, and D. Shih, *ABCDisCo: Automating the ABCD method with machine learning*, arXiv:2007.14400
[3]: V. Mikuni, B. Nachman, and D. Shih, *Online-compatible Unsupervised Non-resonant Anomaly Detection*, arXiv:2111.06417
[4]: E. M. Metodiev, B. Nachman, and J. Thaler, *Classification without labels: Learning from mixed samples in high energy physics*, arXiv:1708.02949
VM, BM, and DN are supported by the U.S. Department of Energy (DOE), Office of Science under contract DE-AC02-05CH11231

@: dnoll@lbl.gov
in: linkedin.com/in/dennis-noll

BERKELEY LAB
Bringing Science Solutions to the World