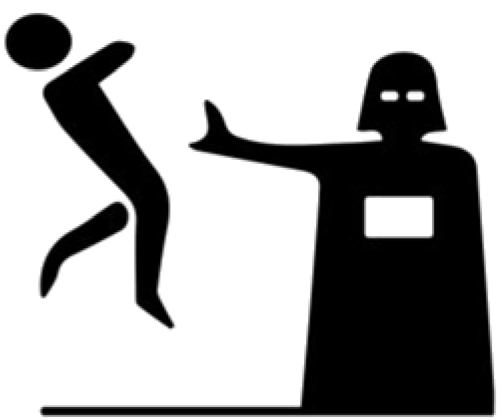
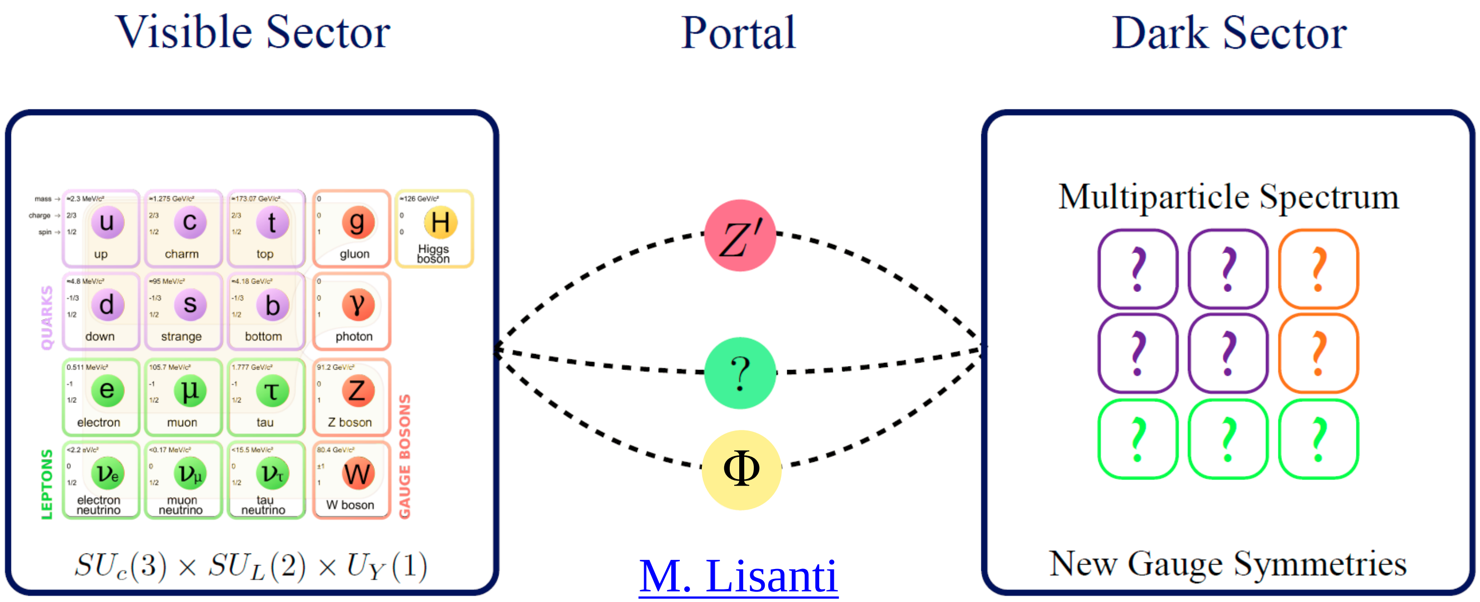
Optimal Mass Variables for Semivisible Jets

Kevin Pedro, Prasanth Shyamsundar
Fermi National Accelerator Laboratory
November 4, 2022



FERMILAB-SLIDES-22-217-PPD-QIS-SCD

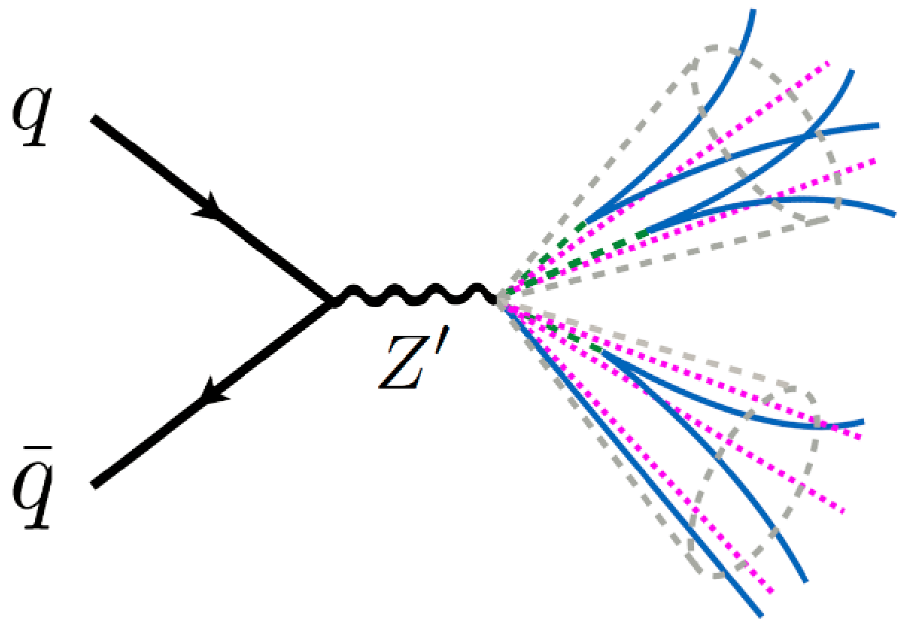
Strongly Coupled Hidden Sectors



- Dark matter may consist of multiple species of composite particles interacting via new, dark forces
  - Visible matter: mostly composite particles, similar density to DM
- In particular, consider a new “dark QCD” force with corresponding dark quarks, dark hadrons, and dark gluons
  - Stable dark hadrons → dark matter candidates!
  - Unstable dark hadrons → decay back to SM, novel phenomenology

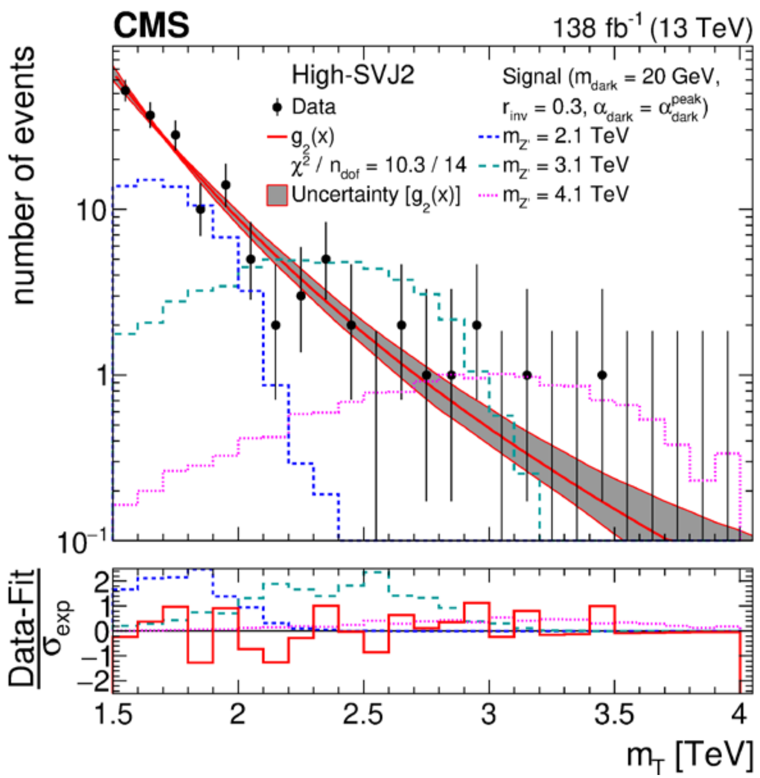
Archetypal Case: *s*-channel

- Massive resonant leptophobic Z'
  - Produces two high-p_T *wide* jets that contain dark matter (missing energy) → semivisible jets! (SVJ)
  - Parameters of interest: $m_{Z'}$, m_{dark} (dark hadron mass), $r_{inv} = \langle N_{stable}/(N_{stable}+N_{unstable})\rangle$

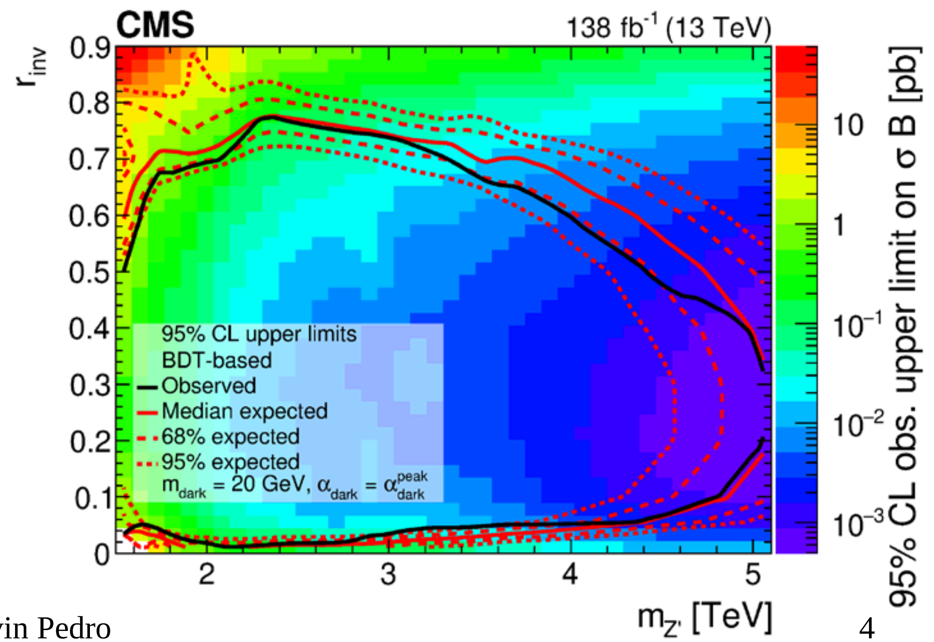


Archetypal Case: *s*-channel

- Massive resonant leptophobic Z'
  - Produces two high-p_T *wide* jets that contain dark matter (missing energy) $\rightarrow$ semivisible jets! (SVJ)
  - Parameters of interest: $m_{Z'}$, m_{dark} (dark hadron mass), $r_{inv} = \langle N_{stable}/(N_{stable}+N_{unstable})\rangle$

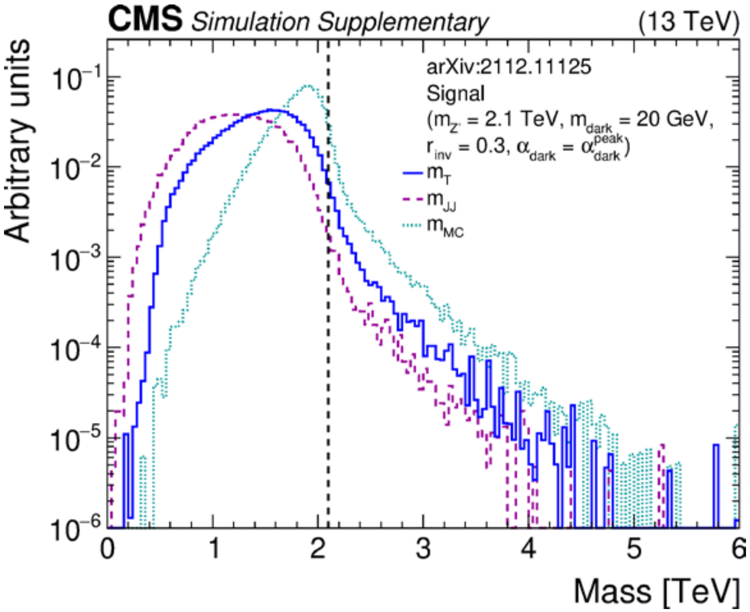


- First search (CMS, [arXiv:2112.11125](https://arxiv.org/abs/2112.11125)):
  - bump hunt in $\mathbf{M_T(JJ,p_T^{miss})}$
  - custom BDT jet tagger



Can we do better?

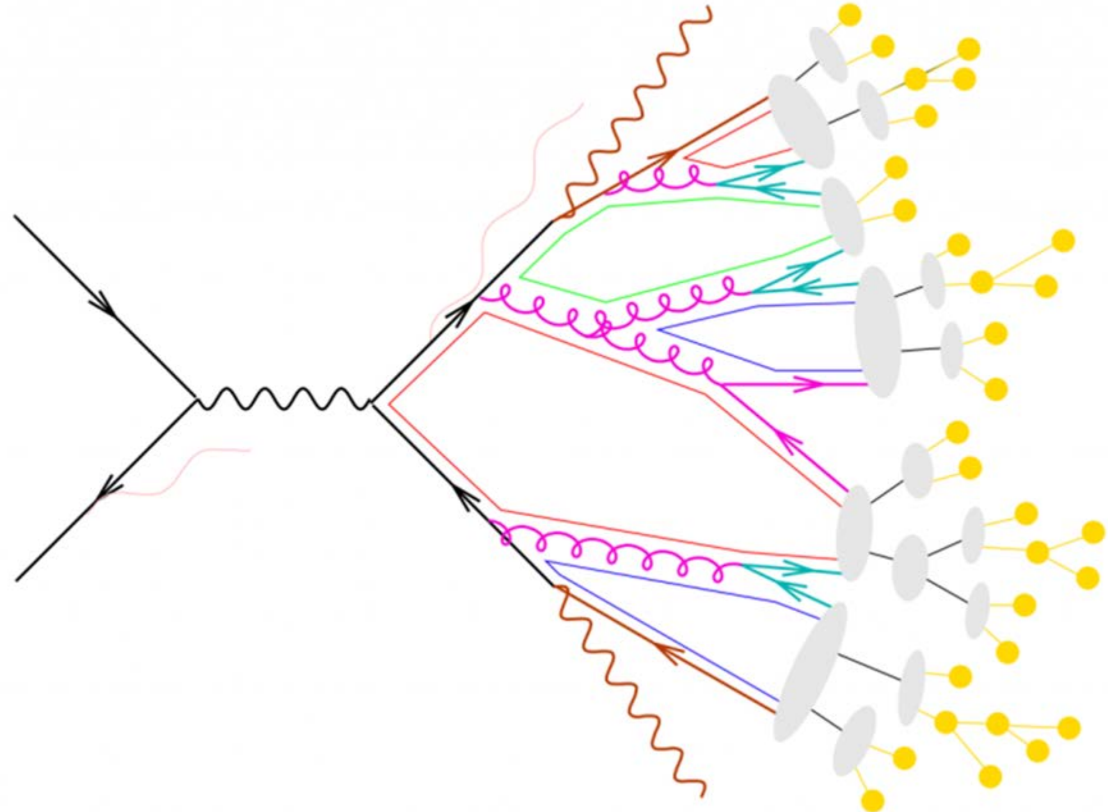
- From this week alone, approximately 100 ideas to improve jet tagger
  - Work in progress
    - Though simple BDT works well w/ adequate feature engineering…
- Today, back to first principles: can we improve Z' reconstruction?
  - M_T resolution is inherently poor
    - Many invisible particles collapsed into a single 2-vector
      - m_{MC} shows “ideal” case w/ full invisible 4-vector
  - Maybe something better is out there…



But can’t you just…?

NO!

- Hadronization is not decay!
  - Electroweak decays conserve parent mass
  - Hadrons form from color connections between quarks & gluons → parton masses *not directly reflected* in final state

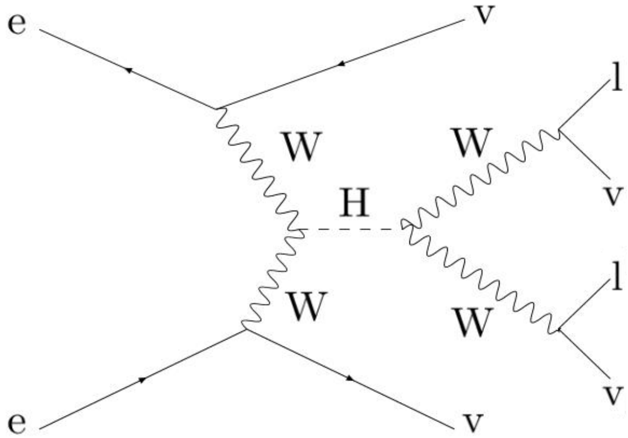


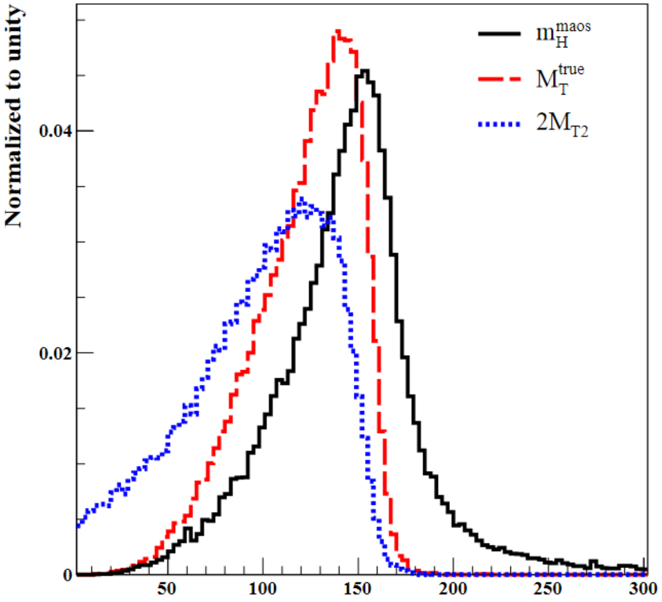
  - Dark QCD even more complicated: unstable dark hadrons decay to $q\bar{q}$
    - Hadronization process takes place again for *each decay*

➢ Existing mass reconstruction techniques focus on easy cases (compared to this)

One Exception

- H → WW: resonance w/ intermediate particles, "semivisible" final state
- M_{T2} Assisted On Shell (MAOS):
  - Use M_{T2} to split p_T^{miss} into two 2-vectors
  - Promote to 4-vectors:
    - Assign mass = 0
    - Assign $q_z = p_z\,(q_T/p_T)$
  - Compute invariant mass
- Works pretty well for Higgs

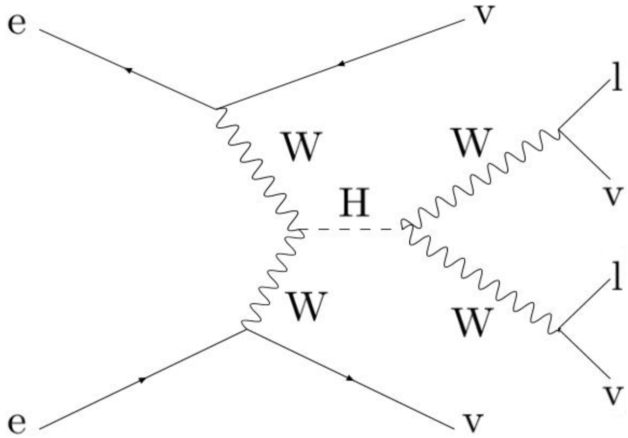


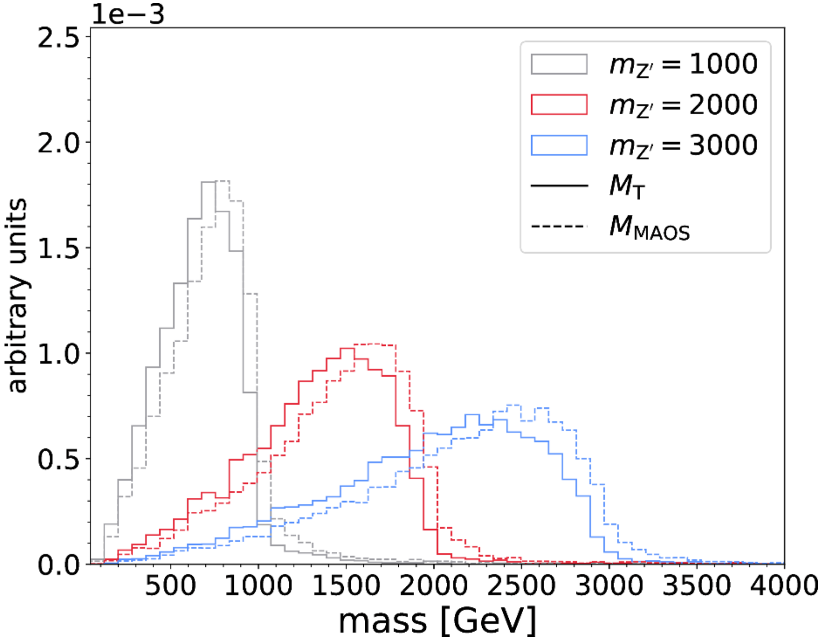


[arXiv:1008.2690](https://arxiv.org/abs/1008.2690)

One Exception

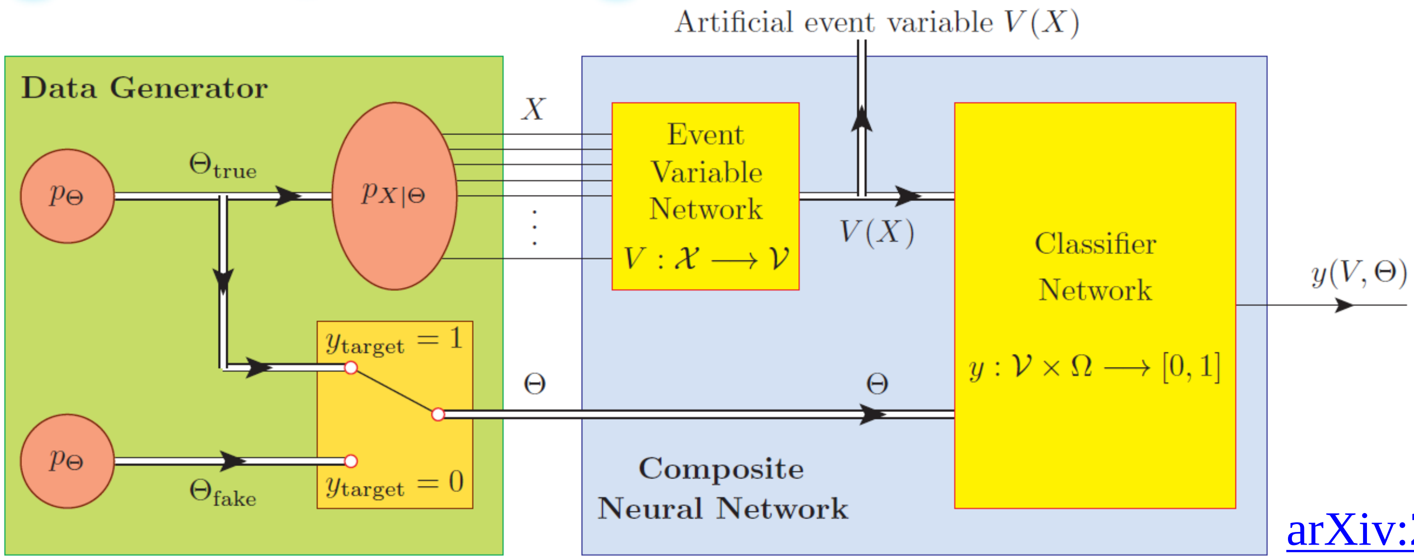
- $H \rightarrow WW$: resonance w/ intermediate particles, "semivisible" final state
- M_{T2} Assisted On Shell (MAOS):
  - o Use M_{T2} to split p_T^{miss} into two 2-vectors
  - o Promote to 4-vectors:
    - ▪ Assign mass = 0
    - ▪ Assign $q_z = p_z\,(q_T/p_T)$
  - o Compute invariant mass
- Works pretty well for Higgs
- Smaller improvement for SVJ
  - o MAOS assumptions not fully valid, e.g. p_T^{miss} not massless





Event Variable Network

Interpretable, semi-supervised machine learning!

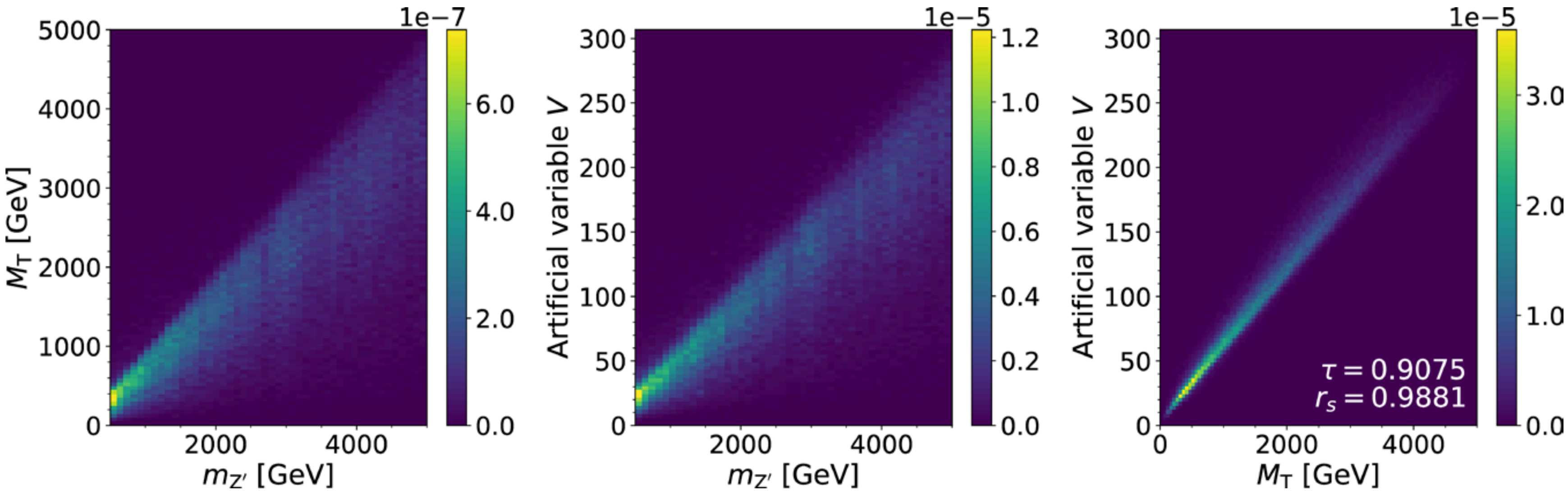


arXiv:2105.10126

- Derive new variable $V(X)$ from inputs X that maximizes *mutual information* with underlying model parameter Θ
  - *Not* a regression: learns an actual, generalized function of inputs
- Both components are simple fully-connected networks (few layers)
  - Classifier uses V (from EVN bottleneck) to distinguish events w/ correct Θ from events w/ wrong Θ (using binary crossentropy)
- Trains in a few minutes on a consumer GPU

Artificial Event Variable for *s*-channel

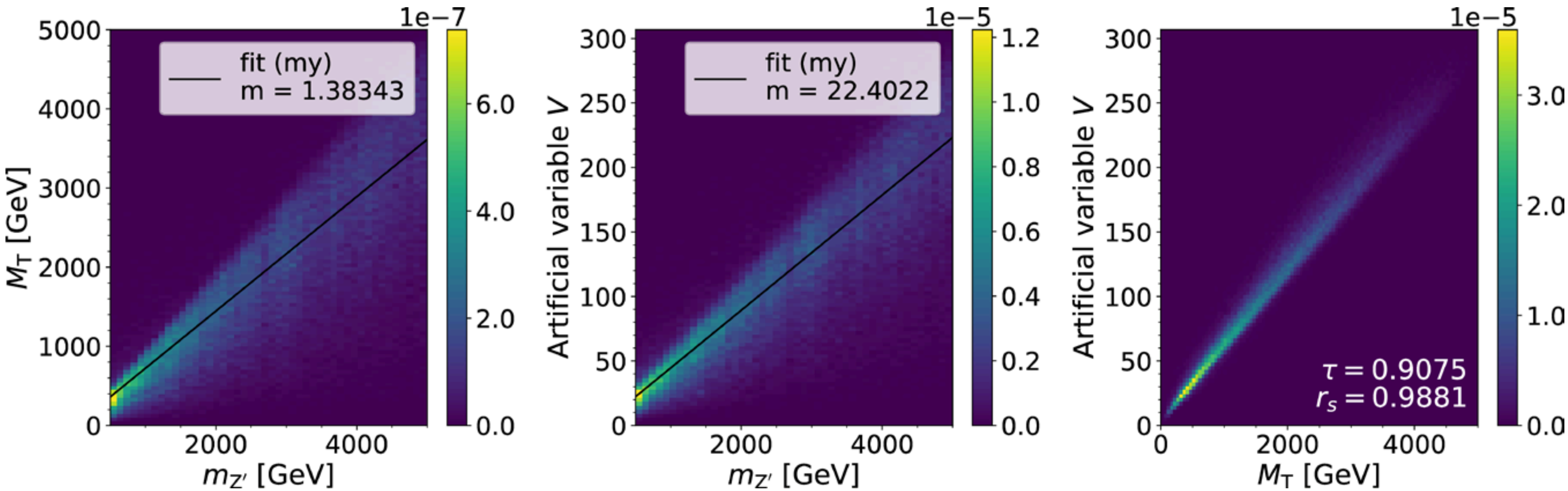
- X = J_1, J_2 4-vectors, p_T^{miss} 2-vector; $\Theta = m_{Z'}$
- Train on signals w/ $m_{Z'}$ = 500…5000 GeV
  - Benchmark values for other parameters: m_{dark} = 20 GeV, r_{inv} = 0.3
  - Generator-level events from Pythia
  - Signal model from CMS search (based on arXiv:1503.00009)



- V correlated w/ M_T, but not identical

Artificial Event Variable for *s*-channel

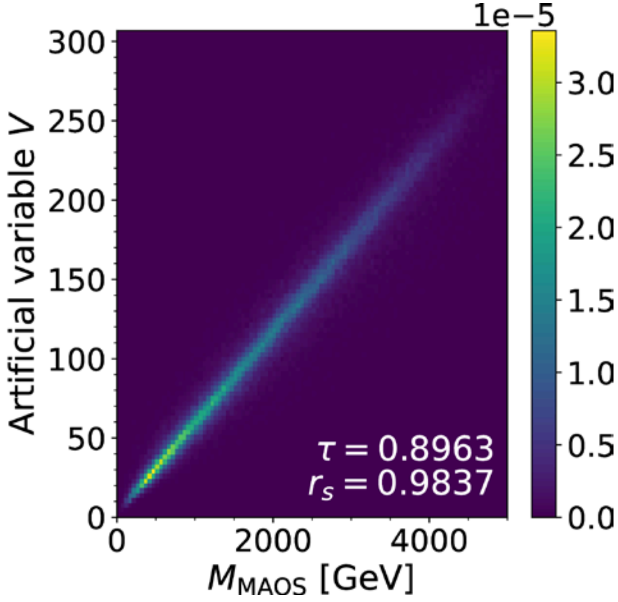
- X = J_1, J_2 4-vectors, p_T^{miss} 2-vector; $\Theta = m_{Z'}$
- Train on signals w/ $m_{Z'}$ = 500…5000 GeV
  - o Benchmark values for other parameters: m_{dark} = 20 GeV, r_{inv} = 0.3
  - o Generator-level events from Pythia
  - o Signal model from CMS search (based on arXiv:1503.00009)

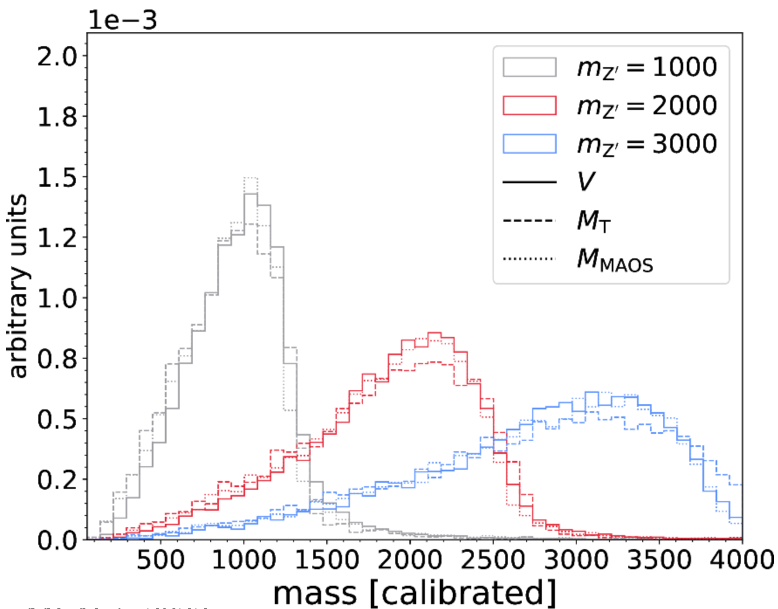


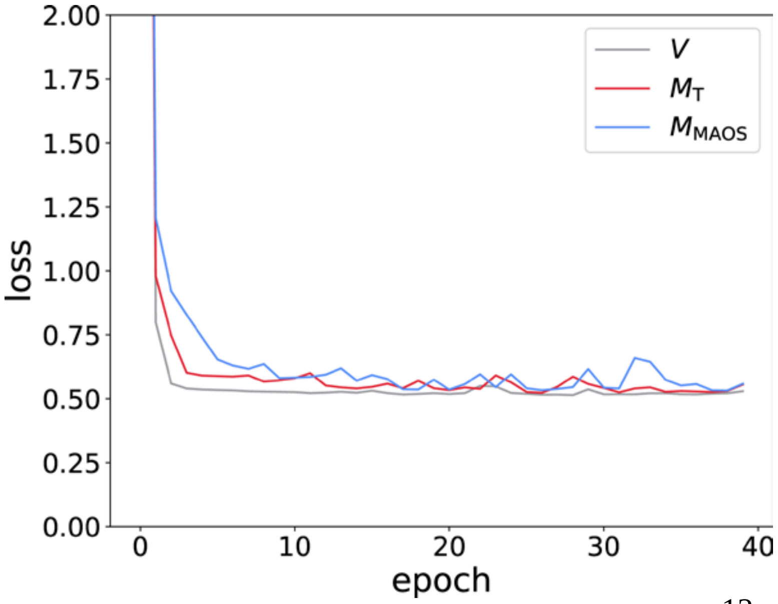
- Note: V has its own (random) scale
- For comparison purposes, calibrate all variables (V, M_T, etc.) to $m_{Z'}$

Investigating *V*

- Correlation w/ M_{MAOS}: close, but not identical
  - Expected from theory: MAOS uses M_{T2}, not a singularity variable
- Looks more like M_{MAOS} than M_T when comparing 1D distributions
- Classifier loss is an estimator of mutual information: apply to each variable
  - Somewhat noisy, but *V* wins

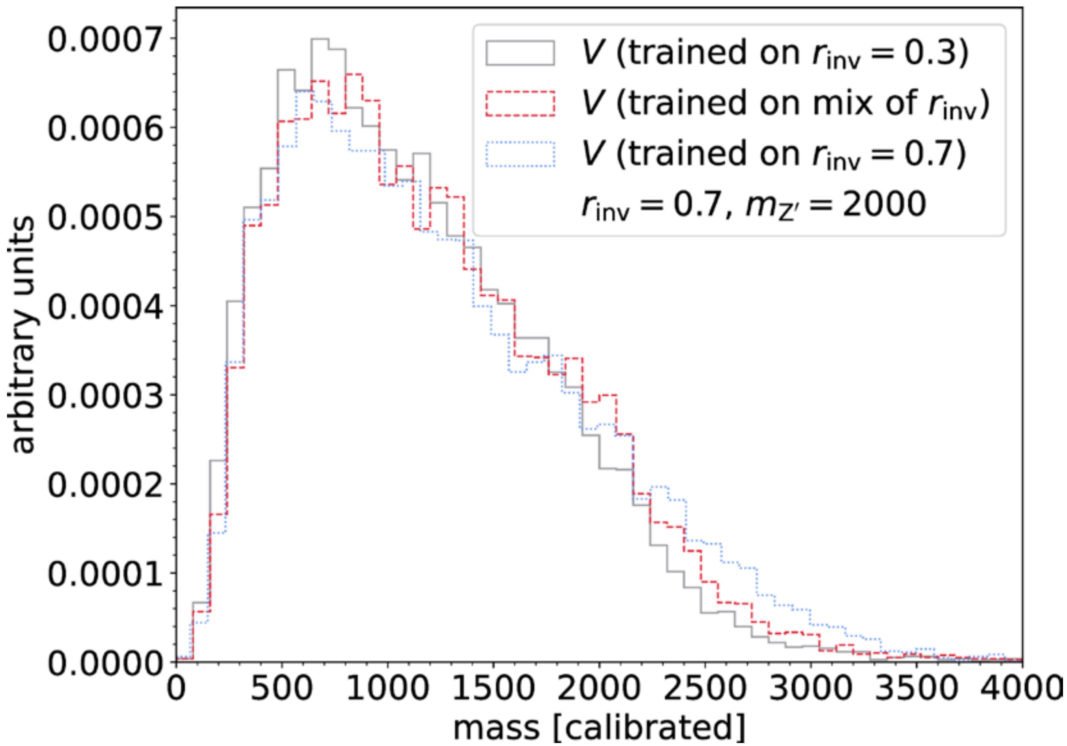






Generalizability

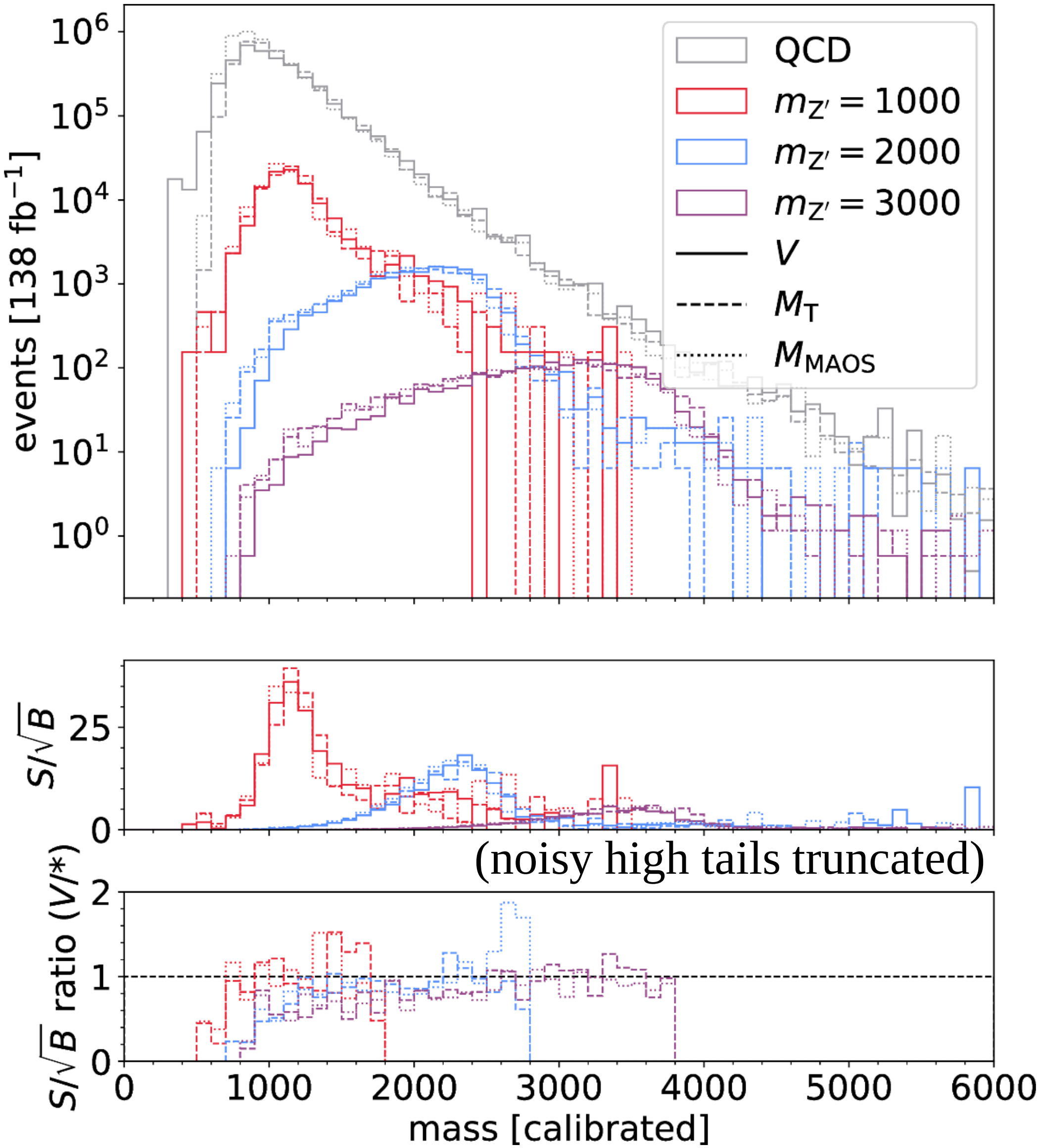
- *V* only trained on signals w/ r_{inv} = 0.3
- Will it work for signals with other values? Yes!
- Very similar results from training on:
  "wrong" r_{inv}, "right" r_{inv}, or mix of r_{inv} values
- Really learning a general mass function!



Analysis Mockup

- Essential ingredients of CMS search:
  - $p_T(J_1,J_2) > 200$ GeV
  - $p_T^{miss}/M_T > 0.15$ (QCD killer, avoids sculpting M_T)
- Just consider generator-level QCD background
  - $p_T^{miss}/\text{mass} > 0.10$
    - Not optimized, just arbitrarily somewhat looser to account for lack of artificial p_T^{miss} without complete detector simulation
    - Applied to each mass variable
- Ignoring trigger, etc.
- Compare bump hunt significance $S/\sqrt{(B)}$

Significance Comparison



Pros:

- *V* gives falling background distribution, as desired!
- ~20% improvement around peak for 2, 3 TeV *Z′*

Cons:

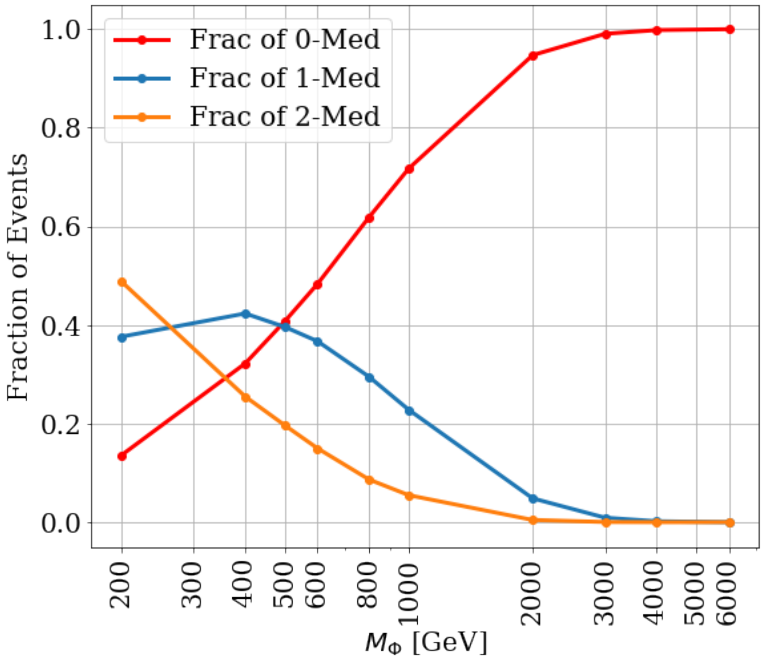
- Not as good for 1 TeV
- Overall minor improvement

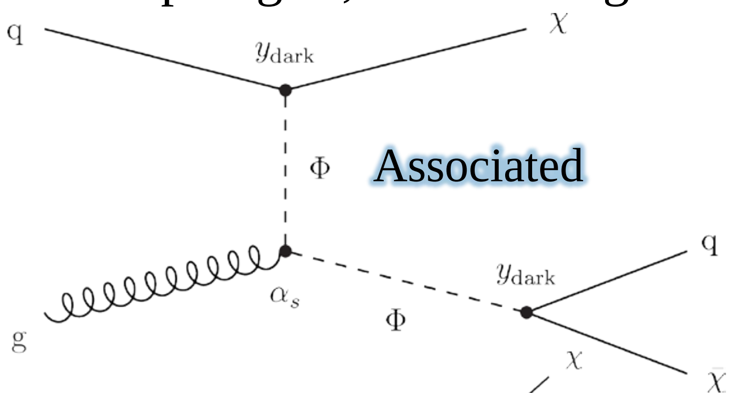
Outlook:

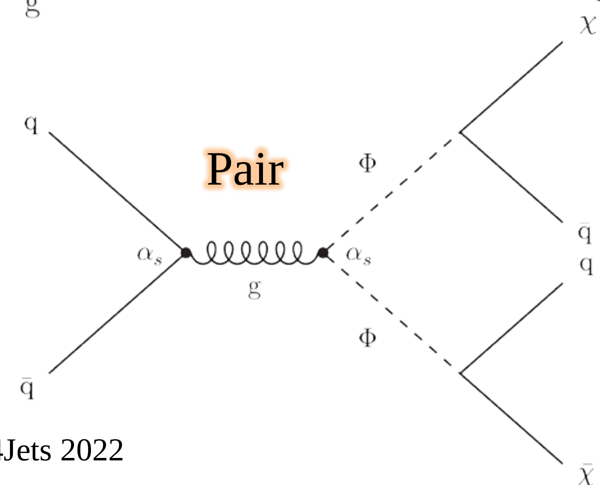
- Optimization (choice of p_T^{miss}/M_T cut) could improve
  - o Could even try to learn optimal QCD killer…
- *V* maybe more useful for measurement than discovery (in this case)

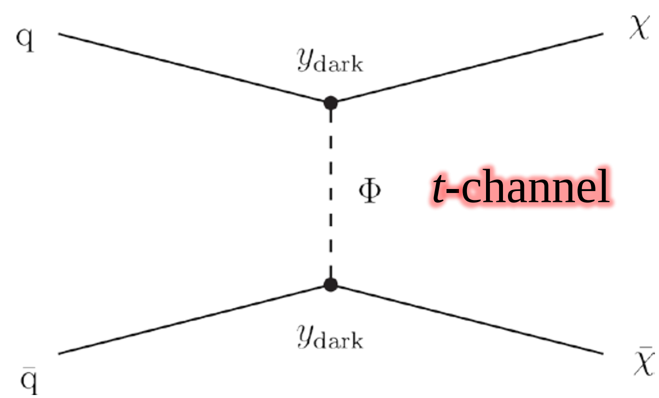
More Cases: *t*-channel

- Bifundamental mediator Φ w/ Yukawa coupling y_{dark} between SM quarks & dark quarks (set to 1)
- Resonant diagrams (left) nontrivial fraction of events up to ~2 TeV
- Novel topologies, less investigated



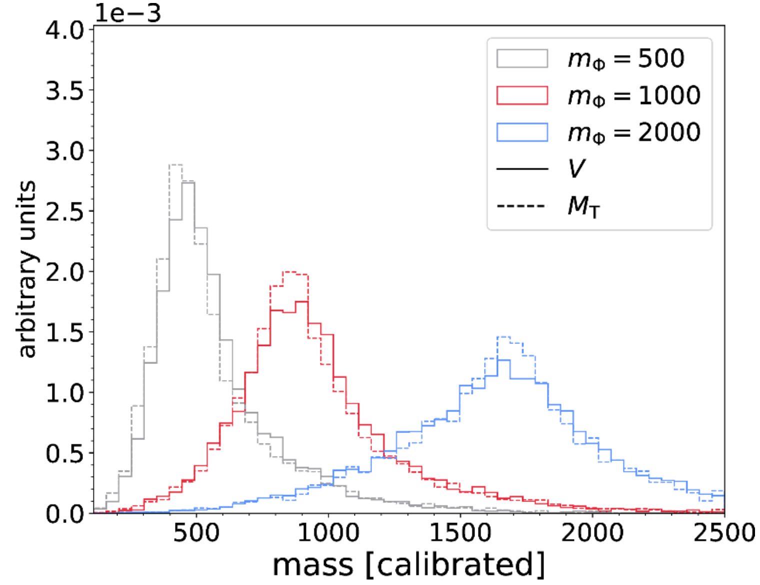


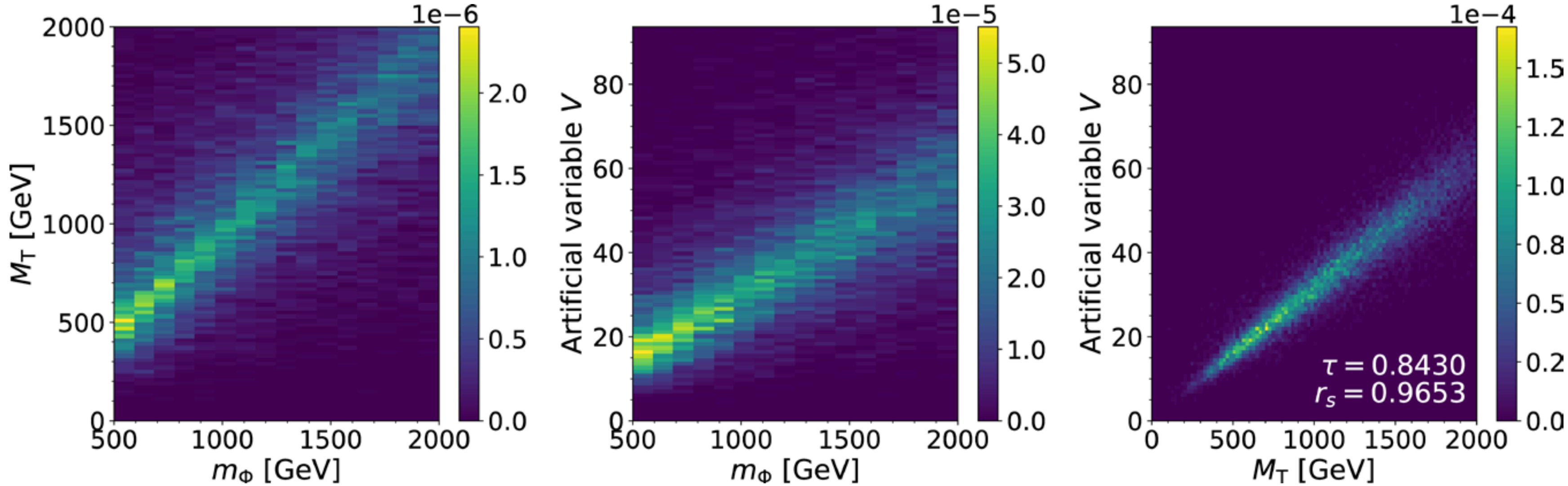




Single Φ Production

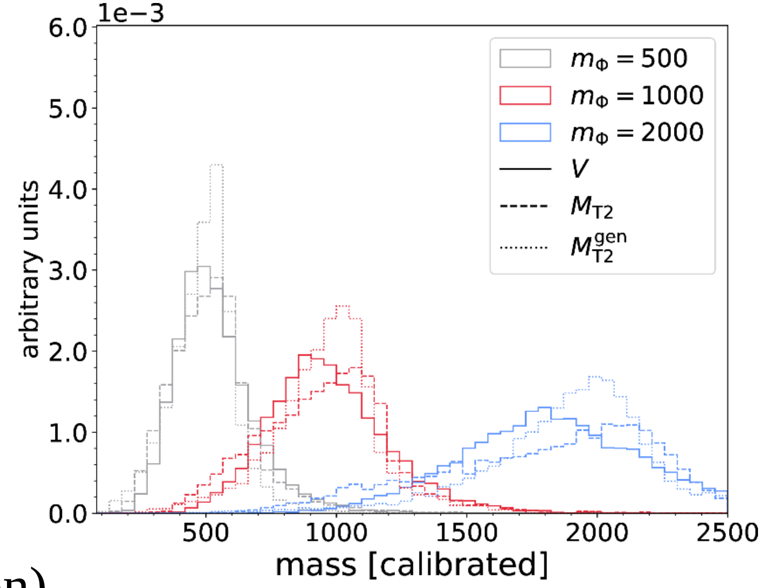
- X = J_1, J_2 4-vectors, p_T^{miss} 2-vector; $\Theta = m_\Phi$
- Could imagine that M_T not optimal because of second (low-p_T) SVJ
- V not 100% correlated with M_T, but looks very similar

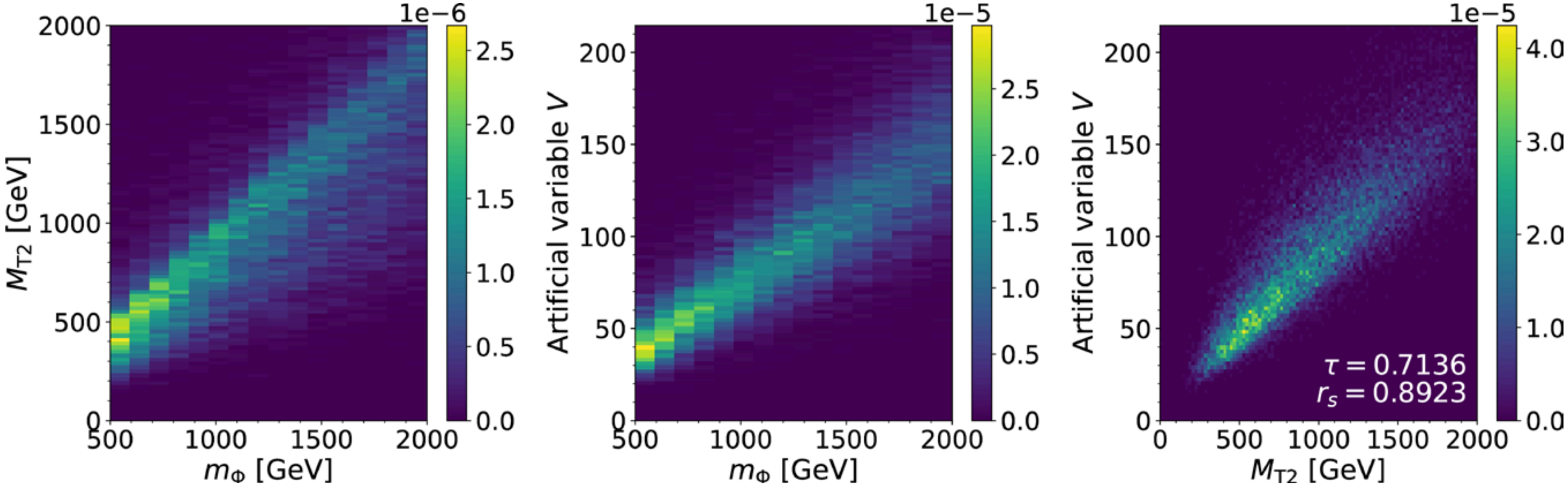




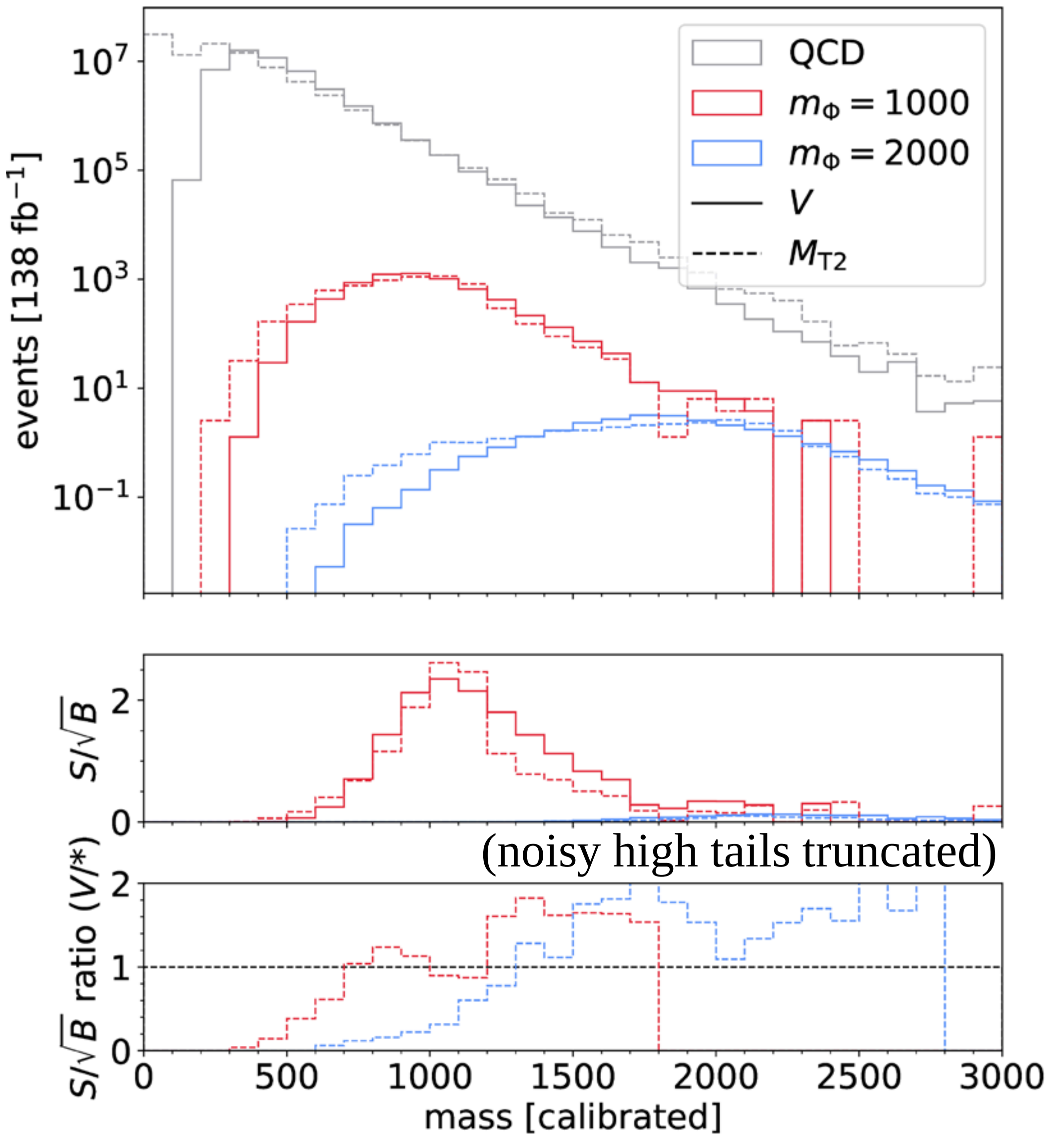
Φ Pair Production

- X = J_1, J_2, J_3, J_4 4-vectors, p_T^{miss} 2-vector; $\Theta = m_\Phi$
- Classical variable here is M_{T2}
  - o Need to do ambiguity resolution: pair jets to minimize ΔM_{JJ}
- V definitely looks better
  - o But not quite as good as M_{T2}^{gen} (using truth-level info for disambiguation)





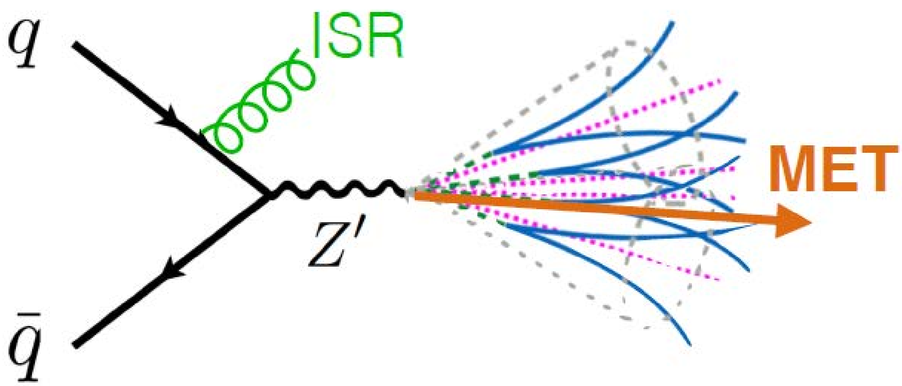
Φ Pair Significance Comparison



- Definite improvement, especially for higher mass
- QCD background shifts toward lower mass
- Promising for a resonant search for Φ mediators
- (see previous caveats about optimization, other improvements, etc.)

Hard Mode: Boosted

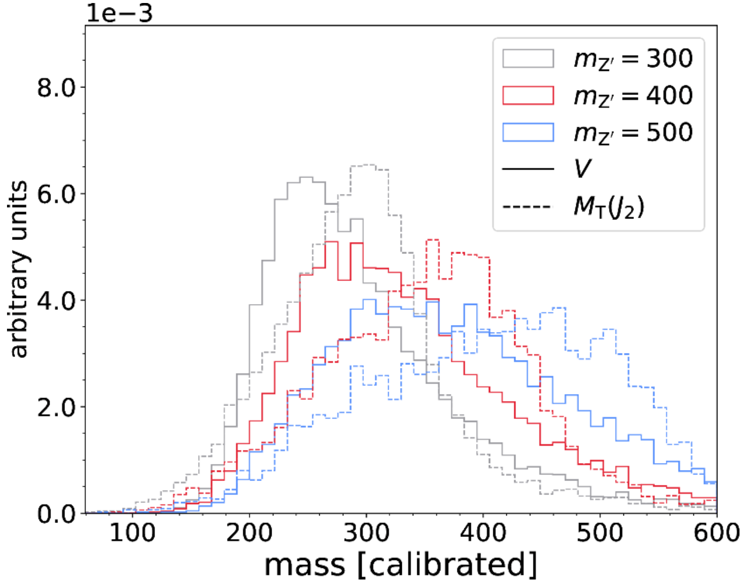
- One *very wide* jet recoiling from hard ISR
- Access low-mass Z' mediators (~250–500 GeV) w/ conventional triggers
  - Requires $p_T(J_1) > \sim 500$ GeV
- $X = J_2$ 4-vectors, p_T^{miss} 2-vector; $\Theta = m_{Z'}$

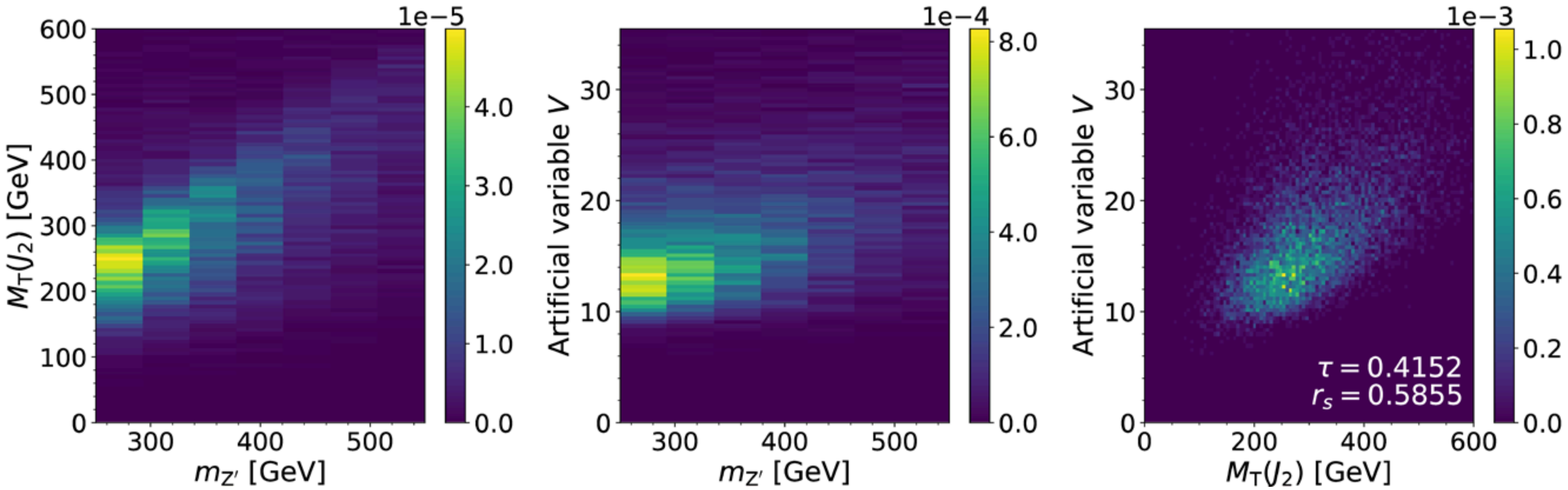


- Generated w/ $p_T(j_1) > 300$ GeV in MadGraph
- Compare to $M_T(J_2, p_T^{miss})$

Boosted Results

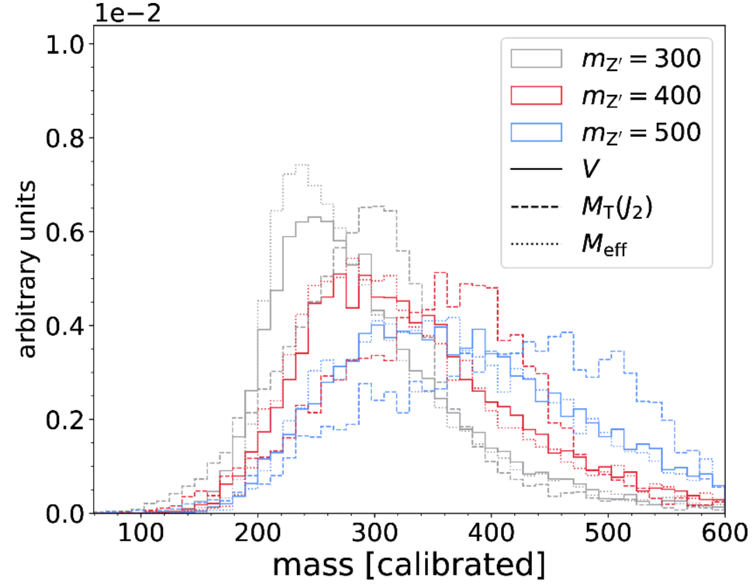
- *V* definitely isn't $M_T(J_2)$…
- But what is it?

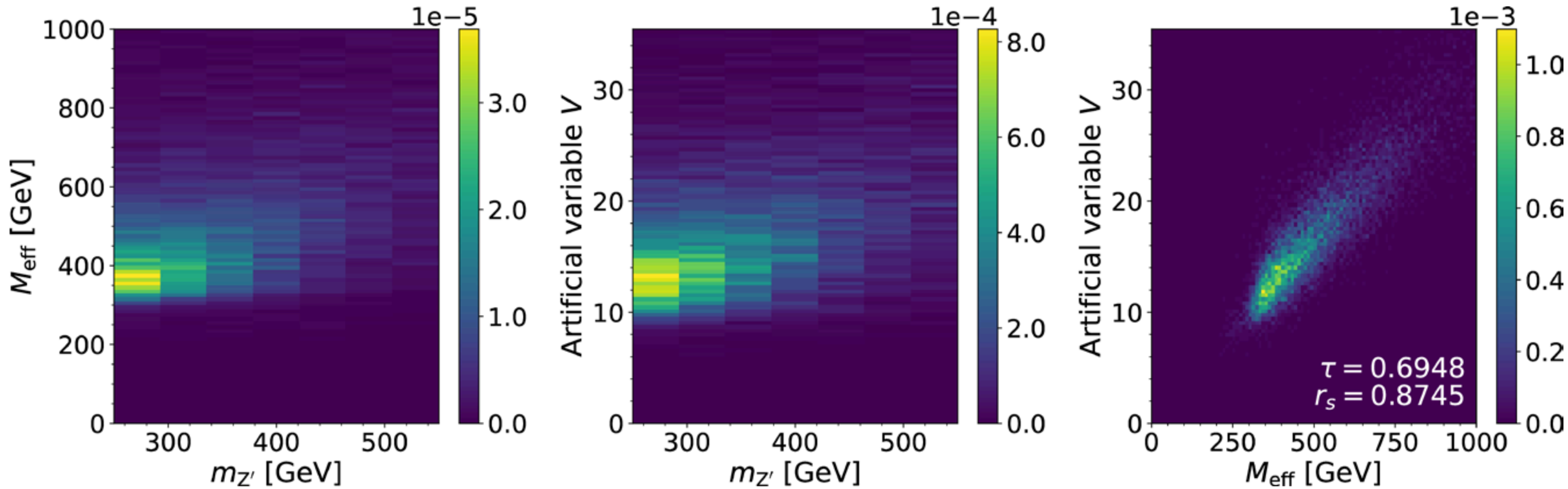




Boosted Results

- *V* definitely isn't $M_T(J_2)$…
- But what is it?
- Something close to $M_{eff} = p_T(J_2) + p_T^{miss}$
- "Narrower", but doesn't discriminate *between* signal models very well

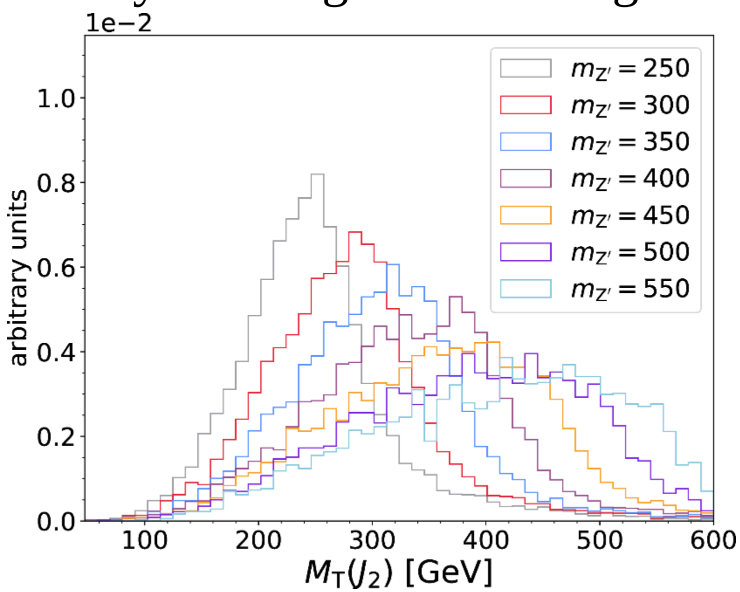




EVN 2.0

See Prasanth's talk from earlier today!

- Instead of comparing correct Θ to totally wrong Θ: compare events w/ "neighboring" Θ values
  - o Force network to improve discrimination between signal models
- Actually use neighbors-of-neighbors: direct neighbors too hard!



0	0	1	0	0	0	0
0	0	0	1	0	0	0
1	0	0	0	1	0	0
0	1	0	0	0	1	0
0	0	1	0	0	0	1
0	0	0	1	0	0	0
0	0	0	0	1	0	0

- Modified loss function: $\left(\varphi(x, \theta_0, \theta_1) = \mathrm{nn}(x, \theta_1) - \mathrm{nn}(x, \theta_0)\right)$

$$L(x_0, \theta_0, x_1, \theta_1) = 0.5 \times \left[L_{\text{logistic}}\big(\varphi(x_0, \theta_0, \theta_1), y = 0\big) + L_{\text{logistic}}\big(\varphi(x_1, \theta_0, \theta_1), y = 1\big)\right]$$

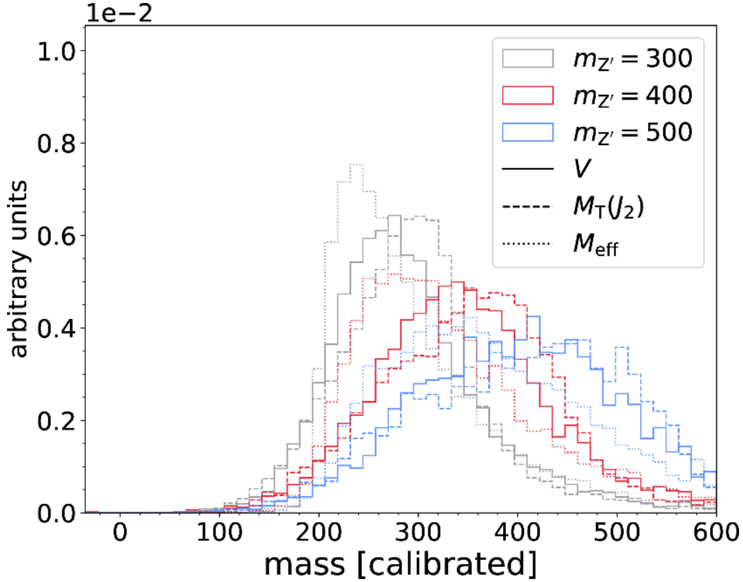
- Needs careful training with hyperparameter optimization (low learning rate, etc.), ~10× epochs vs. EVN 1.0 on easier topologies

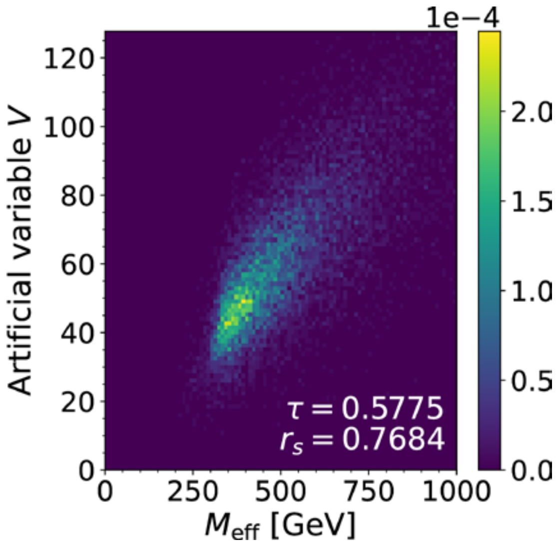
Boosted Results 2.0

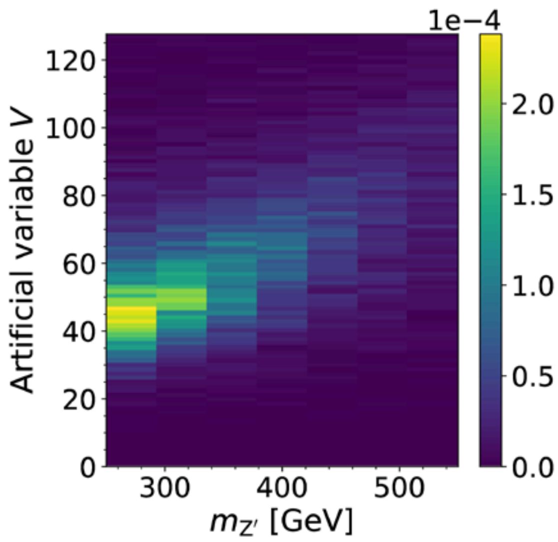
- Now looks more like a hybrid of $M_T(J_2)$ and M_{eff}
- Better signal–signal discrimination
- Closer to M_T at higher masses (harder to boost into a single cone)

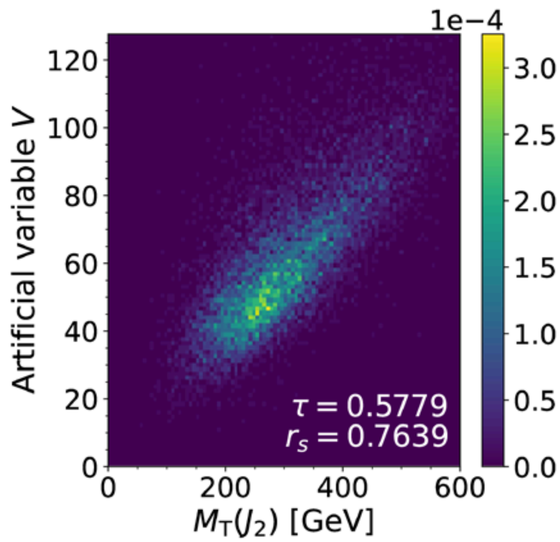
➢ Still room for improvement

  o Continue to refine EVN 2.0



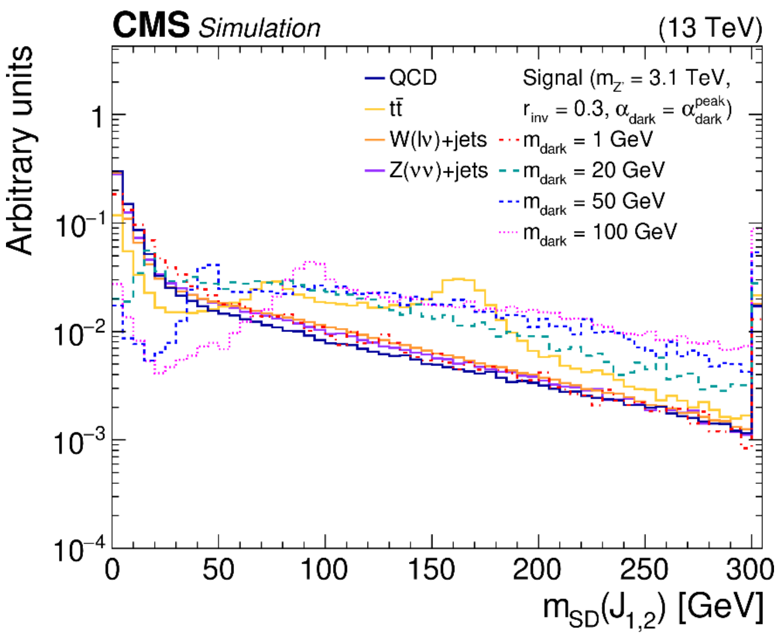




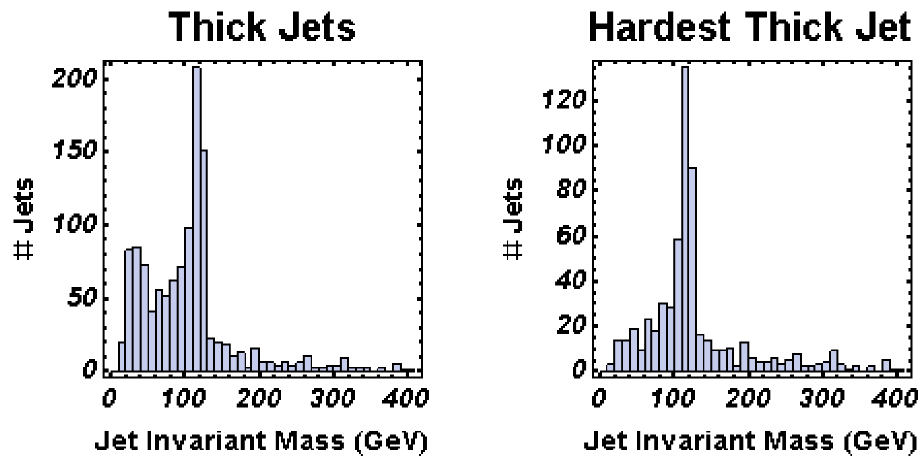


What about m_{dark}?

- Event-level kinematics dominated by $m_{Z'}$ or m_{Φ} (and r_{inv})
- m_{dark} influences jet internal kinematics



Softdrop mass reflects m_{dark}
(hard fragmentation)



Case A2

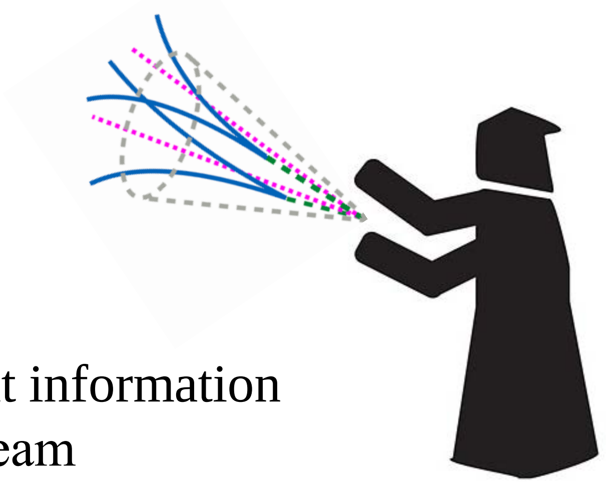
Picking hardest jet with $m_j > 0.15\ p_T$ reflects dark hadron mass (M. Strassler, arXiv:0806.2385)

➢ Can the EVN learn to put dark hadrons back together? Stay tuned!

Conclusion & Outlook

- Applied interpretable, semi-supervised ML to semivisible jet models
  - And learned some new things!
  - Improvement over conventional techniques in some cases
  - Generalizes well: desired behavior on background, out-of-band signals
  - Promising for both searches and, after discovery, measurements
- Next steps:
  - Continue to refine EVN 2.0 for harder cases
  - Look *inside* semivisible jets to learn more
  - Find more applications – many possibilities!
- Bigger question:
  - Can EVN still be interpretable if *unsupervised*?
    - Learn appropriately-sized bottleneck for event information *and* provide comprehensible output → the dream
- Let's continue to think about how to make ML work *for us* (not vice versa)
  - And may the dark force be with us all!

Preprint coming soon!



Acknowledgments

- Co-author of this study: Prasanth Shyamsundar
- Other authors of arXiv:2105.10126: Doojin Kim, Kyoungchul Kong, Konstantin T. Matchev, Myeonghun Park
- Useful discussions: Tim Cohen, Sandeep Madireddy, Siddharth Mishra-Sharma, Matt Strassler
- CMS collaborators: Eshwen Bhal, Florencia Canelli, Cesare Cazzaniga, Annapaola De Cosa, Florian Eble, Daniel Elvira, Sarah Eno, Colin Fallon, Henning Flacher, Aran Garcia-Bellido, Thomas Klijnsma, Benjamin Krikler, Chris Madrid, Steve Mrenna, Sara Nabili, Jeremi Niedziela, Giorgia Rauco, Roberto Seidita, Chin Lung Tan
- Support: Fermi National Accelerator Laboratory, managed and operated by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy

Backup

Software Versions

- Python: 3.6.9
- numpy: 1.19.5
- scipy: 1.5.4
- matplotlib: 3.3.4
- tensorflow: 2.6.0
- uproot: 4.1.3
- awkward: 1.5.0
- vector: 0.8.4
- magiconfig: 2.3.1

Hyperparameter settings

Parameter	Z'	Φ	$\Phi\Phi$	Boosted	Boosted 2.0
% train	64	64	64	64	64
% validation	16	16	16	16	16
% test	20	20	20	20	20
Random seed	10	10	10	10	10
Batch size	5000	5000	5000	5000	100
Epochs	40	40	40	40	100+200
AEV nodes	128, 64, 64, 64, 32	128, 64, 64, 64, 32	128, 64, 64, 64, 32	128, 64, 64, 64, 32	16, 8, 4
AEV activation	ReLU	ReLU	ReLU	ReLU	ReLU
Aux nodes	16, 16, 16	16, 16, 16	16, 16, 16	16, 16, 16	16, 16, 16
Aux activation	ReLU	ReLU	ReLU	ReLU	ReLU
Learning rate	0.001	0.001	0.001	0.001	0.0001+0.001
Optimizer	Adam	Adam	Adam	Adam	Adam

Generator details

- QCD: Pythia 8.2, $\hat{p}_T$ = [15, 7000] GeV (unflattened), ~10M events
- Z′: Pythia 8.2, $m_{Z'}$ = [500, 5000] GeV (steps of 100 GeV)
  - r_{inv} = 0.3: ~12K events per $m_{Z'}$ value
  - r_{inv} = 0.1, 0.5, 0.7: ~6K events per $m_{Z'}$ value
  - Test samples w/ $m_{Z'}$ = [1000, 2000, 3000] GeV: ~10K events each
- Φ: MadGraph 2.6.5 + Pythia 8.2, m_Φ = [500, 2000] GeV (steps of 100 GeV), ~10K events per m_Φ value (train & test samples)
- ΦΦ: MadGraph 2.6.5 + Pythia 8.2, m_Φ = [500, 2000] GeV (steps of 100 GeV), ~10K events per m_Φ value (train & test samples)
- Boosted Z′: MadGraph 2.6.5 + Pythia 8.2, $m_{Z'}$ = [250, 550] GeV (steps of 50 GeV), $p_T(j_1)$ > 300 GeV, ~50K events per $m_{Z'}$ value (train & test samples)
  - Require Z′, χ_1, χ_2 (dark quarks) to have $\Delta R < 1.5$ w/ J_2 (truth level)

- MadGraph settings based on https://github.com/smsharma/SemivisibleJets

More on SVJ

Presentation of CMS results: Semivisible Jet Workshop

- Many plots & references therein
- Select references reproduced here

Theory:

- M. J. Strassler and K. M. Zurek, “Echoes of a hidden valley at hadron colliders”, Phys. Lett. B 651 (2007) 374, arXiv:hep-ph/0604261.
- T. Cohen, M. Lisanti, and H. K. Lou, “Semivisible jets: Dark matter undercover at the LHC”, Phys. Rev. Lett. 115 (2015) 171804, arXiv:1503.00009.
- T. Cohen, M. Lisanti, H. K. Lou, and S. Mishra-Sharma, “LHC searches for dark sector showers”, JHEP 11 (2017) 196, arXiv:1707.05326.
- G. Albouy et al., “Theory, phenomenology, and experimental avenues for dark showers: a Snowmass 2021 report”, arXiv:2203.09503.

Tagging:

- M. Park and M. Zhang, “Tagging a jet from a dark sector with jet-substructures at colliders”, Phys. Rev. D 100 (2019) 115009, arXiv:1712.09279.
- E. Bernreuther, T. Finke, F. Kahlhoefer, M. Krämer, and A. Mück, “Casting a graph net to catch dark showers”, SciPost Phys. 10 (2021) 046, arXiv:2006.08639.
- F. Canelli, A. de Cosa, L. Le Pottier, J. Niedziela, K. Pedro, M. Pierini, “Autoencoders for Semivisible Jet Detection”, JHEP 02 (2022) 074, arXiv:2112.02864.