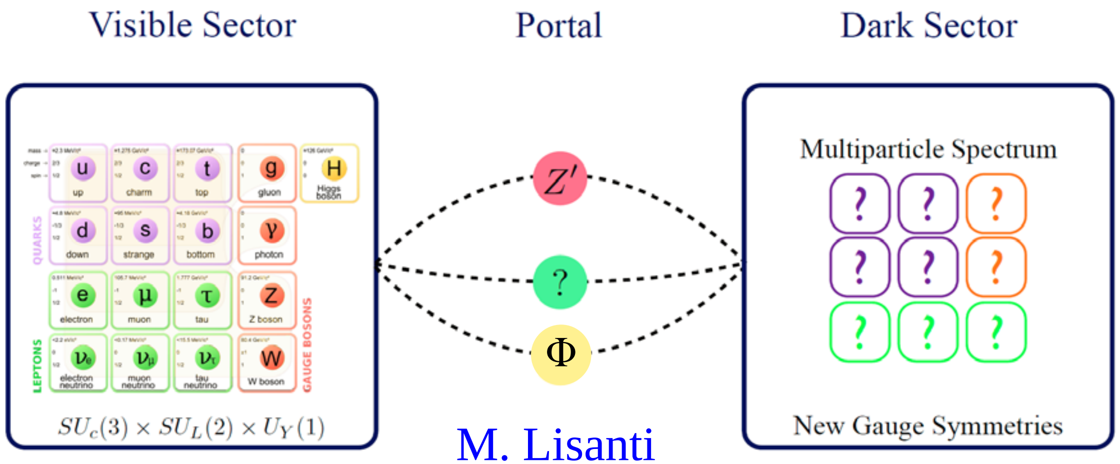
Searching for Strongly Coupled Dark Sectors with Unsupervised and Generative Learning

Kevin Pedro
(Fermilab)

August 11, 2024

FERMILAB-SLIDES-24-0160-CMS-CSAID-PPD

Strongly Coupled Dark Sectors



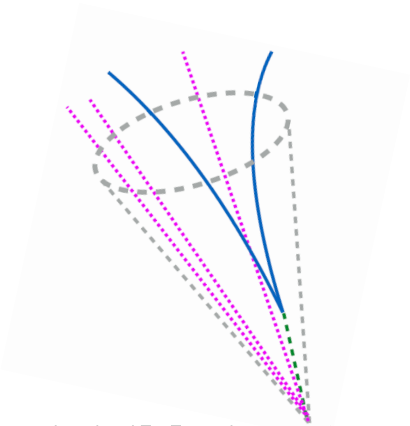
M. Lisanti

- Numerous sources of evidence that dark matter exists and behaves differently from visible matter
- But no direct experimental evidence of its nature
- What if DM is *hiding* in existing collider data?

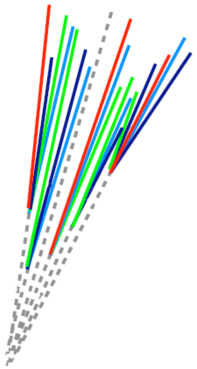
- Dark sector may consist of multiple species of particles interacting via new "dark QCD" force
- Analogous to SM: dark quarks, dark gluons, dark hadrons (stable and unstable)

- DM abundance arises from asymmetry mechanism → no annihilation
- DM interactions with ordinary matter highly suppressed → no direct detection

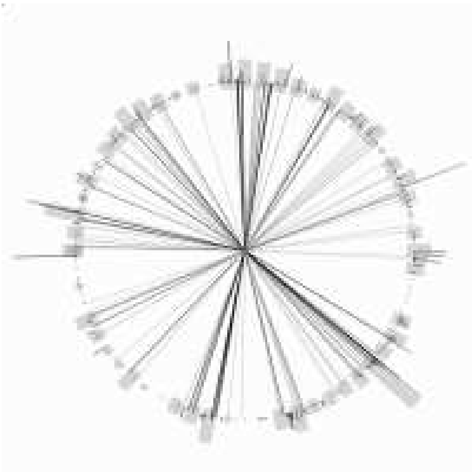
Phenomenology



Semivisible jets (SVJs)
mixture of stable and unstable dark hadrons
→ p_T^{miss} aligned with jets



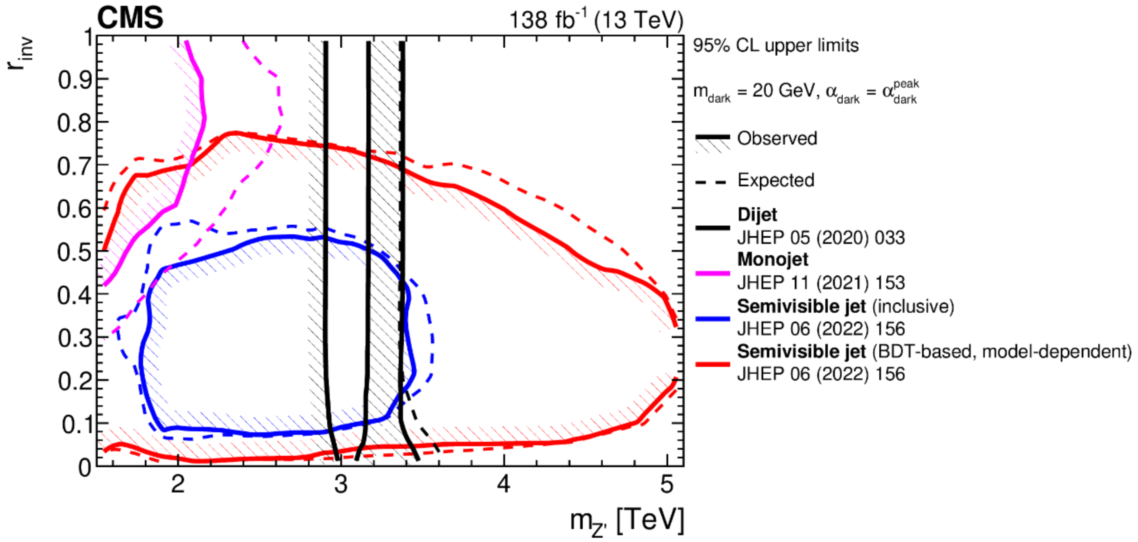
Emerging jets (EMJs)
dark hadrons decay after some lifetime
→ multiple displaced vertices within each jet

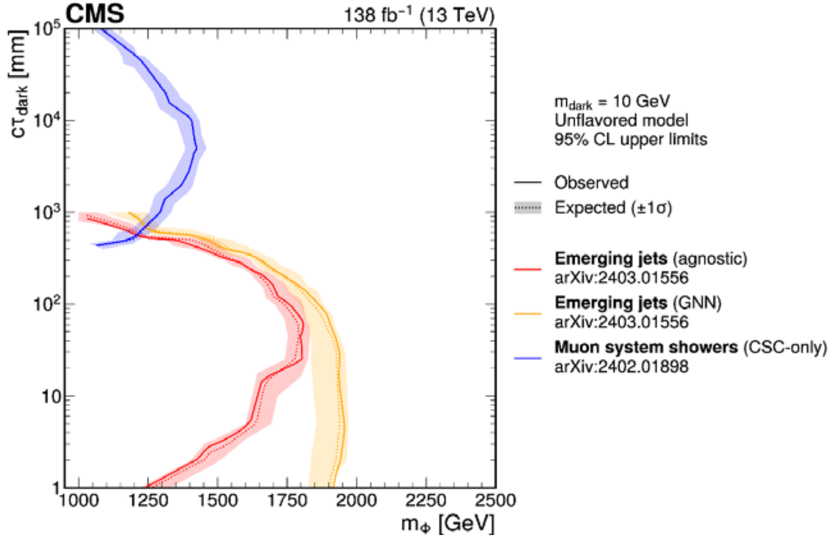


Soft unclustered energy patterns (SUEPs)
confining theory with large ‘t Hooft coupling, beyond QCD-like regime
→ spherical distribution of low-p_T tracks

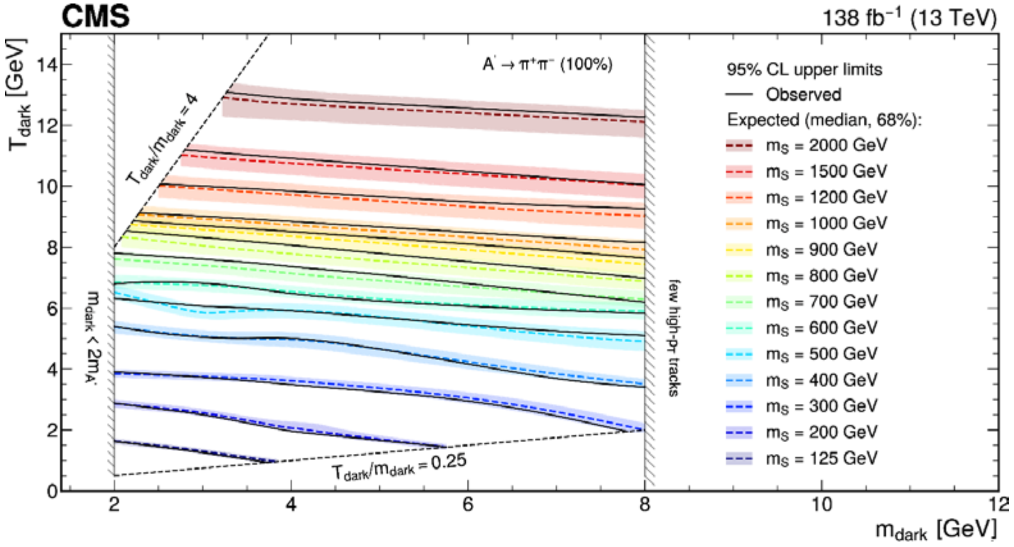
Latest Results

- CMS has executed *first searches* for all dark QCD phenomena using LHC Run 2 data





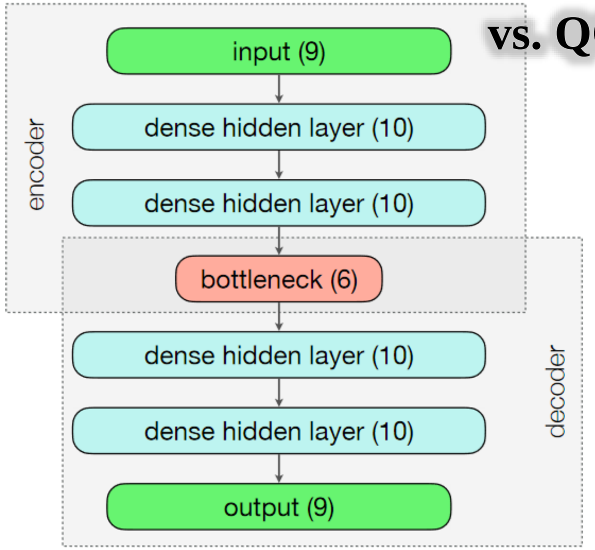
- *Complementarity* between existing strategies and new, dedicated searches



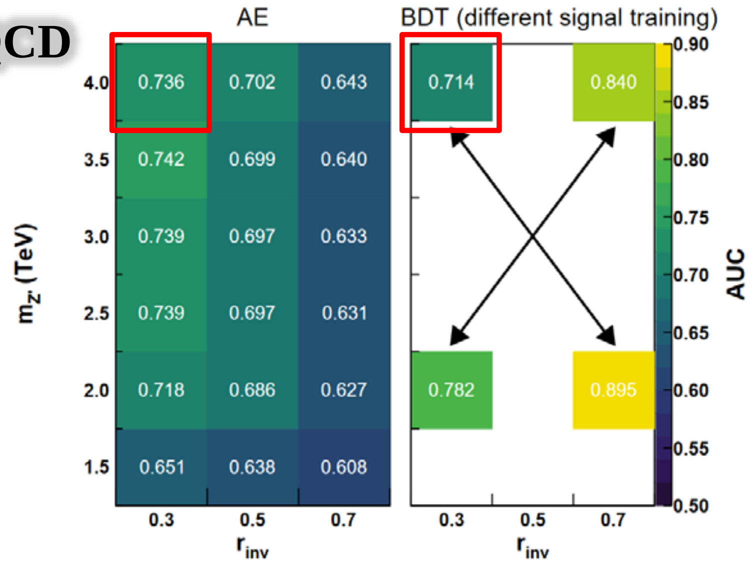
- *Supervised ML* to "tag" new kinds of jets:
  - Significantly increases sensitivity
  - Cost: model dependence

Autoencoders for Semivisible Jets

arXiv:2112.02864



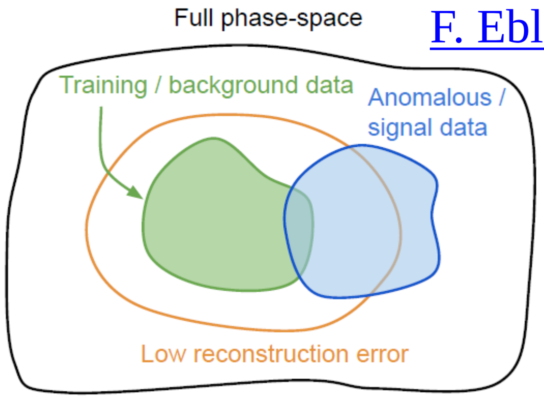
vs. QCD



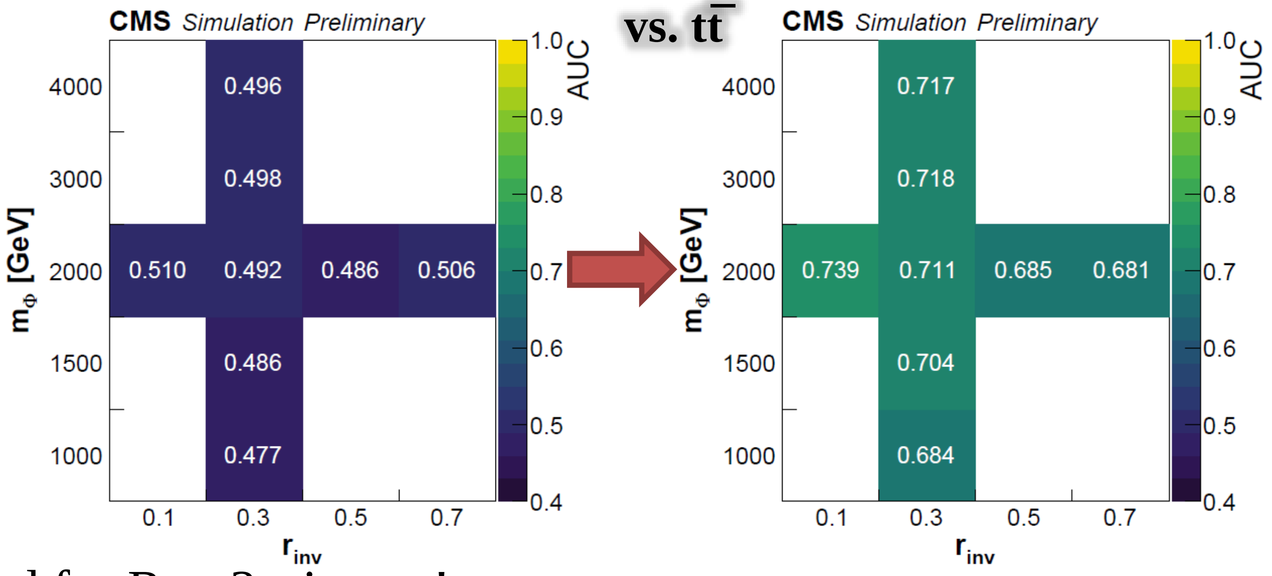
- Learn latent representation that can accurately reconstruct *background*
  - o Signal not used in training; identified later via high *reconstruction error*
- Autoencoder can *outperform* BDT on signals with different parameter values
- Works well against QCD background, but not $t\bar{t}$ … *complexity bias*

- Normalized autoencoder formalism:
  - o *Sample* from low-error space during training to prevent over-generalizing
- Use Energy Mover's Distance to choose best model (prevent mode collapse)

F. Eble



vs. $t\bar{t}$

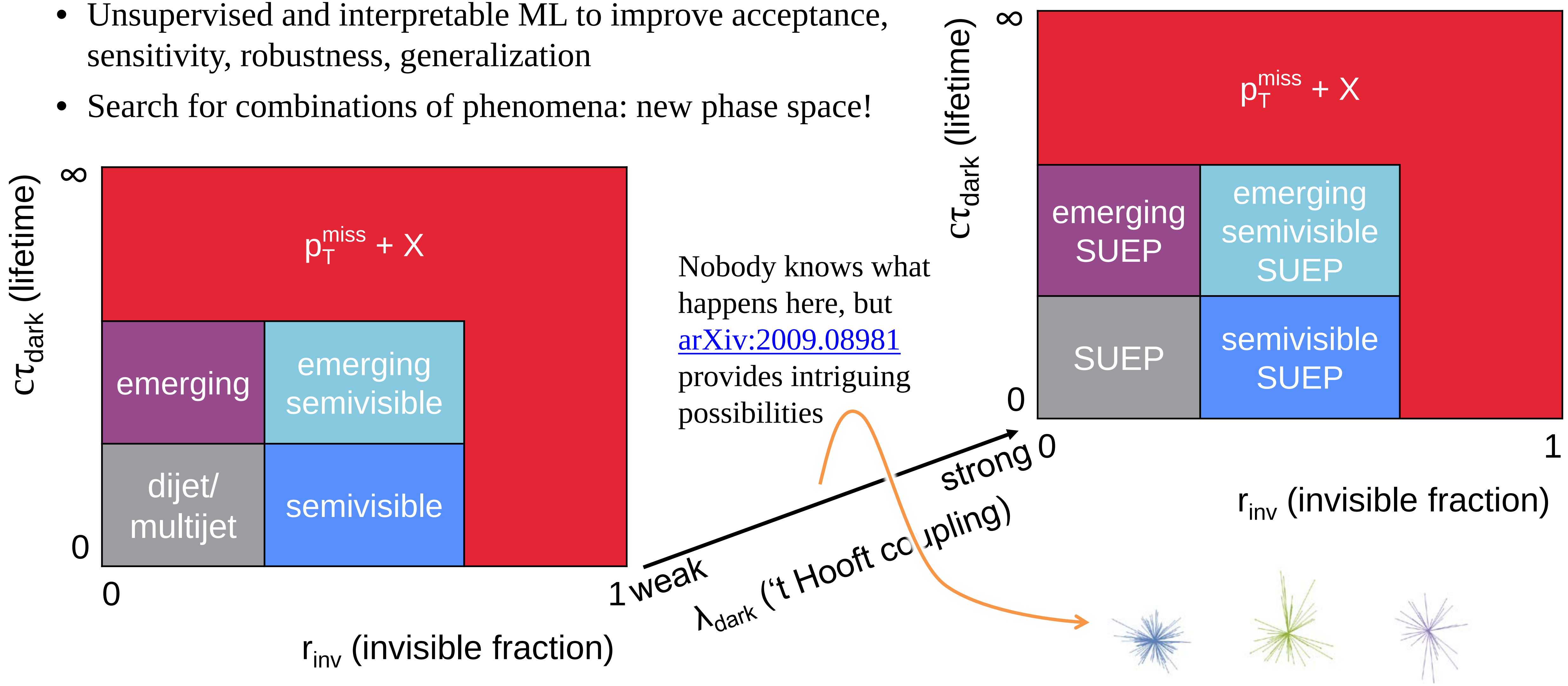


➢ Now being pursued for Run 3 triggers!

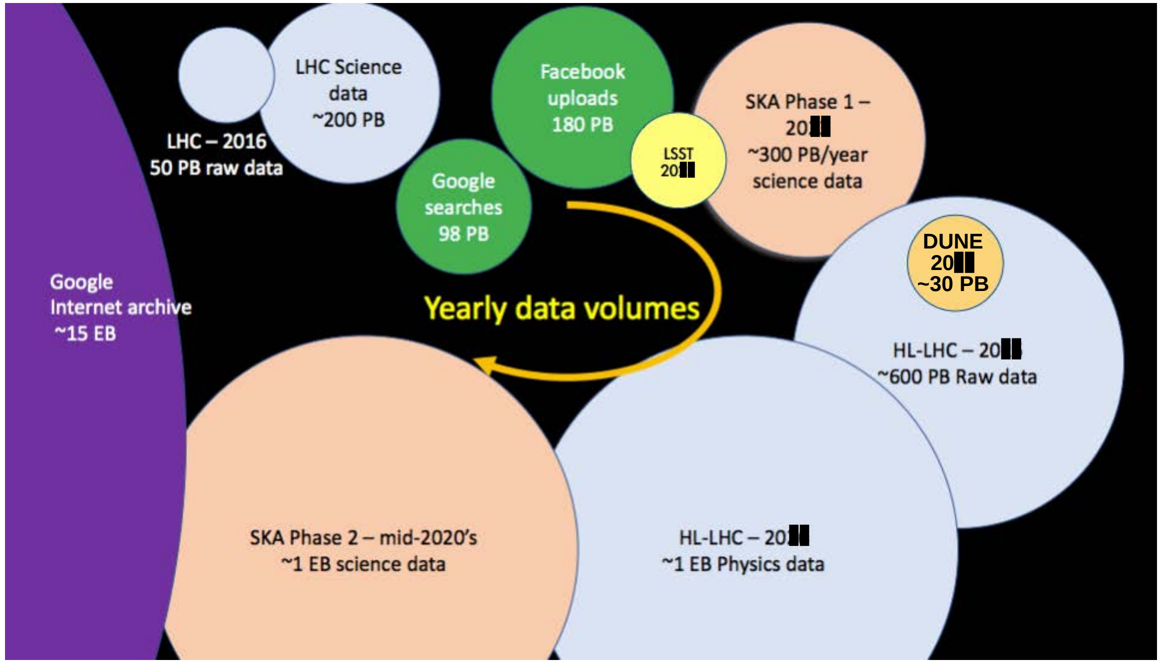
The Future of Dark QCD Searches

- Expand to more production modes (vector, scalar, bifundamental, Higgs, …)
- Unsupervised and interpretable ML to improve acceptance, sensitivity, robustness, generalization
- Search for combinations of phenomena: new phase space!

Nobody knows what happens here, but [arXiv:2009.08981](https://arxiv.org/abs/2009.08981) provides intriguing possibilities

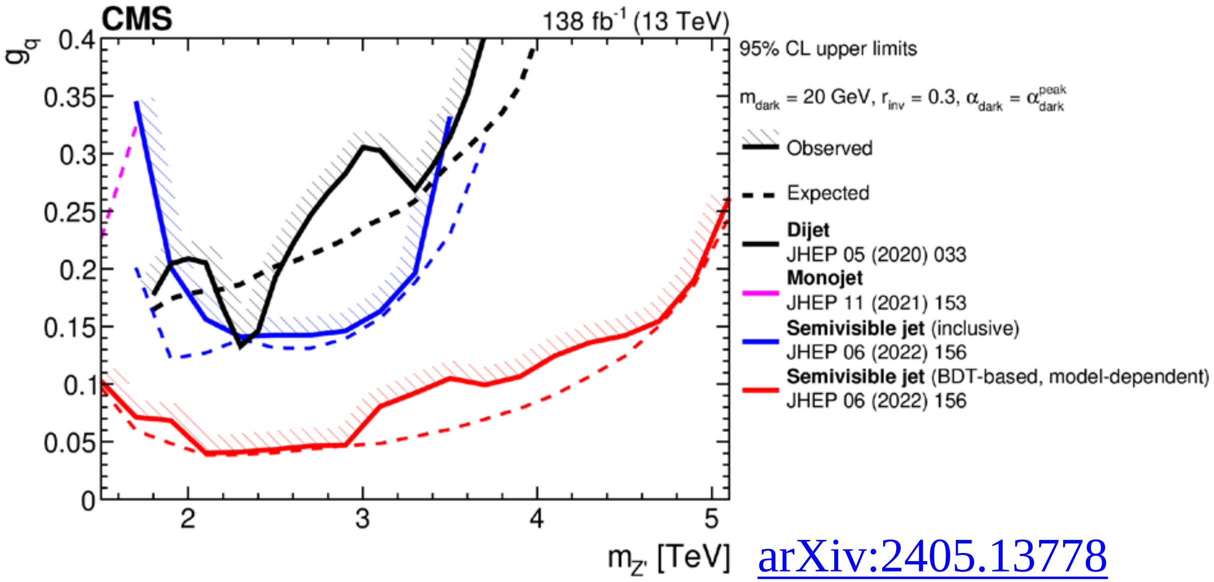


High-Luminosity Future



- *Challenges* to process, store, and analyze this upcoming flood of data
- Simulation is critical for analysis design and interpretation
  - o Will need to deliver more events with more complexity and more accuracy… while using *smaller fraction* of computing

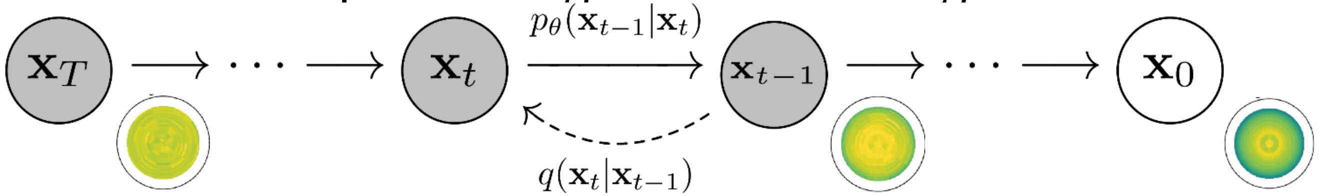
- A new precision era is imminent: HL-LHC, DUNE, LSST, SKA
  - o 10× or more data vs. existing experiments
- *Opportunities* for most precise measurements and discovery of rare processes
  - o Dark QCD w/ very weakly coupled mediators?
    - Models have dozens of parameters…



arXiv:2405.13778

ML For Simulation

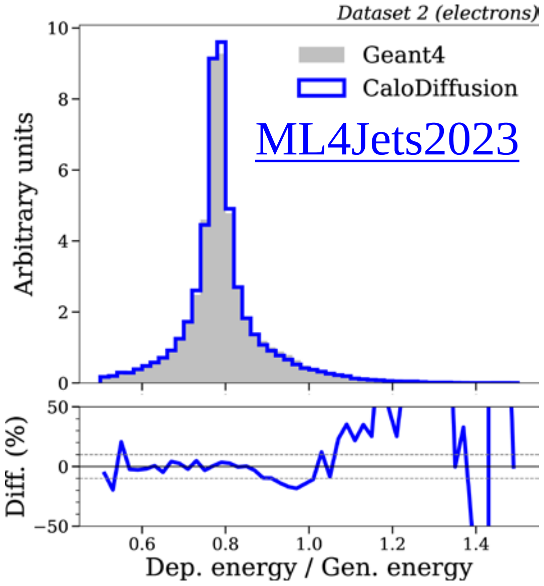
- CaloDiffusion ([arXiv:2308.03876]): current state-of-the-art model
  - o Adapt industry generative techniques to irregular calorimeter geometries



- Now the standard for comparison and a platform for further development: [arXiv:2401.13162], [arXiv:2406.12898]

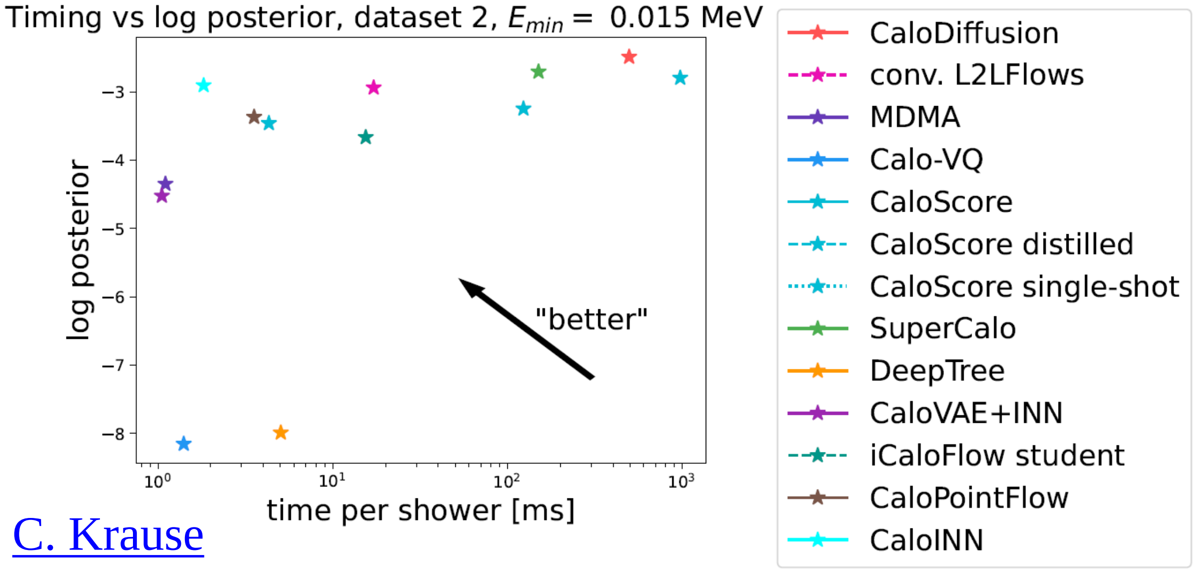
- Excellent agreement with full simulation even in difficult global quantities

➢ Most accurate entry in community challenge



[ML4Jets2023]

- Also in development: ML refinement of high-level variables
  - o Correct for any remaining inaccuracies
  - o "Perturbative" approach to detector simulation

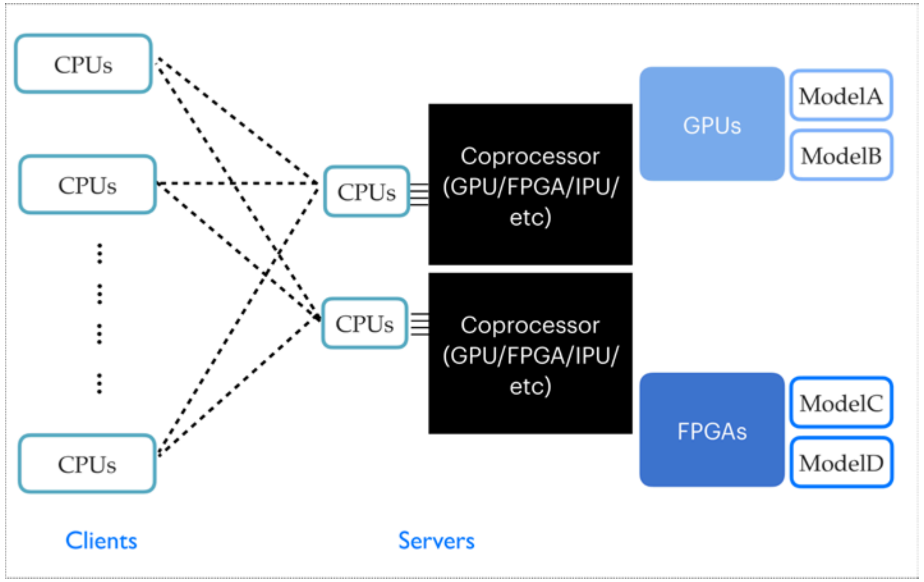


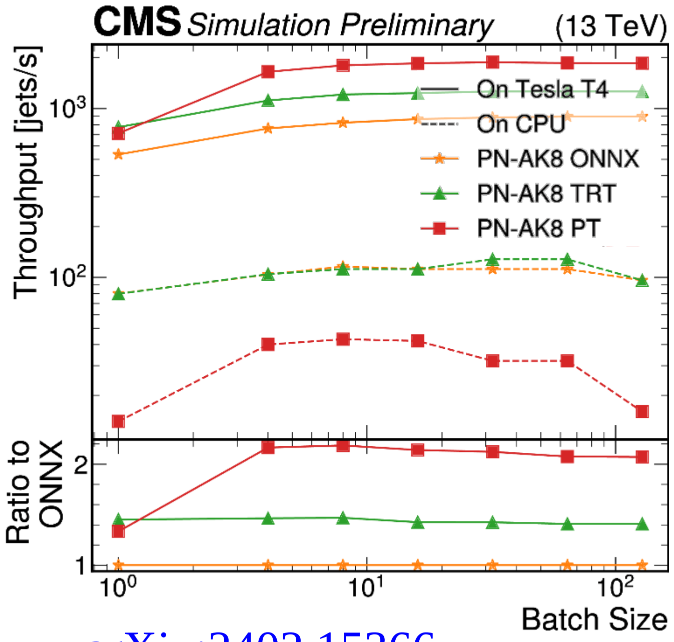
[C. Krause]

Computing for ML

- ML algorithms use a restricted set of operations (mostly matrix multiplications)
  - o Natural and easy to accelerate on specialized coprocessors
- *Most flexible* approach: inference as a service
  - o Abstract away specific computing elements: client makes request, server delivers
    - ▪ Use CPUs, GPUs, FPGAs, IPUs… with no code changes!
  - o Example: ~10× speedup in ParticleNet on GPU vs. CPU
    - ▪ Algorithm latency becomes essentially *invisible* with asynchronous calls in offline processing
    - ▪ Can batch *across events* for optimal GPU utilization → maximize throughput
  - o Similar 10× speedup in CaloDiffusion

➢ ML can *solve* high-luminosity challenges
  - o Result: *better, smarter, faster* physics

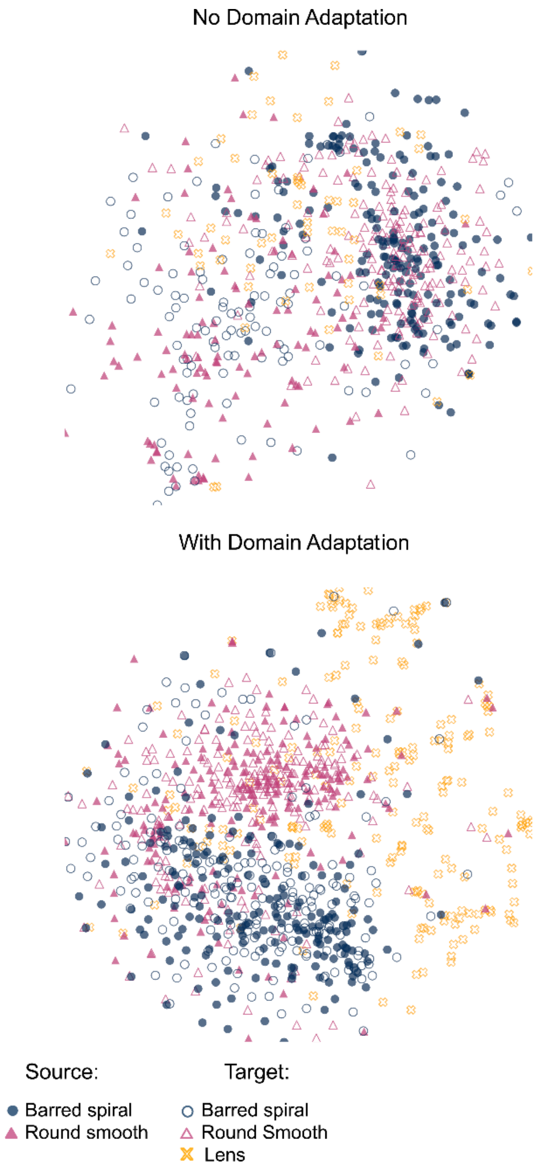




arXiv:2402.15366

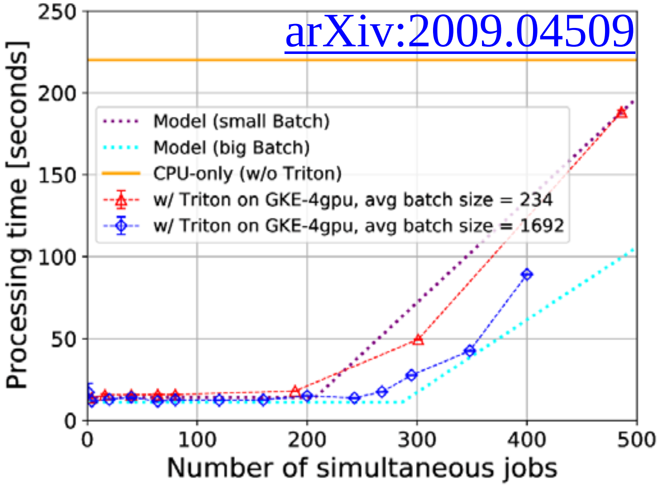
Beyond Colliders

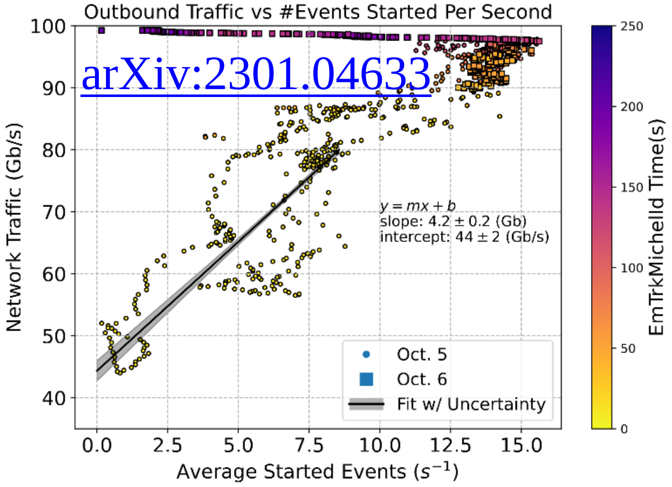
- Astrophysics:
  - o Goal: classify events (e.g. galaxy mergers)
  - o Multiple data sources, limited-accuracy simulation
    - ➢ Use semi-supervised *contrastive learning* (adaptive clustering + entropy separation)
  - o Domain adaptation improves performance and robustness
  - o Physically meaningful latent space
  - o Anomaly detection for unknown classes



arXiv:2302.02005

Neutrinos:
- o Inference as a service vital for large image classification algorithms
- o Algorithm level: 17× speedup using GPU
  - ▪ Analytical model of GPU saturation effects
- o Datacenter level: reprocess entire protoDUNE dataset in weeks instead of months
  - ▪ Mitigate network saturation via concurrency limits





Acknowledgments

- Everything shown here would not be possible without many amazing collaborators! Thanks to:
  - URA for this award and for their support of Fermilab's science missions
  - Fermilab Particle Physics Directorate and Computational Science & AI Directorate
  - US CMS Collaboration and LHC Physics Center
  - International CMS collaboration at CERN
  - Dark QCD theorists, colleagues at other labs
  - Funders: DOE Energy Frontier and Computational HEP, USCMS Computing Operations, NSF E-CAS
  - Industry partners: Nvidia, Microsoft, Google, Graphcore
  - My family, especially my wife, Titas

2024 US CMS Annual Collaboration Meeting (Princeton)

Backup

Publications

Dark QCD:
- CMS Collaboration, "Search for resonant production of strongly-coupled dark matter in proton-proton collisions at 13 TeV", JHEP 06 (2022) 156, arXiv:2112.11125.
- 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.
- G. Albouy et al., "Theory, phenomenology, and experimental avenues for dark showers: a Snowmass 2021 report", Eur. Phys. J. C 82 (2022) 1132, arXiv:2203.09503.
- K. Pedro and P. Shyamsundar, "Optimal Mass Variables for Semivisible Jets", SciPost Phys. Core 6 (2023) 067, arXiv:2303.16253.
- F. Eble, "Unsupervised tagging of semivisible jets with normalized autoencoders in CMS", PoS EPS-HEP2023 (2024) 491.
- CMS Collaboration, "Search for new physics with emerging jets in proton-proton collisions at $\sqrt{s}$ = 13 TeV", arXiv:2403.01556, submitted to JHEP.
- CMS Collaboration, "Search for soft unclustered energy patterns in proton-proton collisions at 13 TeV", arXiv:2403.05311, submitted to Phys. Rev. Lett.
- CMS Collaboration, "Dark sector searches with the CMS experiment", arXiv:2405.13778, submitted to Phys. Rept.

ML/computing:
- J. Duarte et al., "FPGA-accelerated machine learning inference as a service for particle physics computing", Comp. Soft. Big Sci. 3 (2019) 13, arXiv:1904.08986.
- K. Pedro, "Current and Future Performance of the CMS Simulation", Eur. Phys. J. Web Conf. 214 (2019) 02036.
- K. Pedro, "Integration and Performance of New Technologies in the CMS Simulation", Eur. Phys. J. Web Conf. 245 (2020) 02020, arXiv:2004.02327.
- D. Rankin et al., "FPGAs-as-a-Service Toolkit (FaaST)", Proc. H2RC (2020) 38, arXiv:2010.08556.
- G. Amadio et al., "GeantV: Results from the prototype of concurrent vector particle transport simulation in HEP", Comp. Soft. Big Sci. 5 (2021) 3, arXiv:2005.00949.
- M. Wang, T. Yang, et al., "GPU-accelerated machine learning inference as a service for computing in neutrino experiments", Front. Big Data 3 (2021) 604083, arXiv:2009.04509.
- J. Krupa, K. Lin, et al., "GPU coprocessors as a service for deep learning inference in high energy physics", Mach. Learn. Sci. Tech. 2 (2021) 035005, arXiv:2007.10359.
- A. Ćiprijanović et al., "DeepAdversaries: Examining the Robustness of Deep Learning Models for Galaxy Morphology Classification", Mach. Learn. Sci. Tech. 3 (2022) 035007, arXiv:2112.14299.
- S. Banerjee et al., "Denoising Convolutional Networks to Accelerate Detector Simulation", J. Phys. Conf. Ser. 2438 (2023) 012079, arXiv:2202.05320.
- A. Adelmann et al., "New directions for surrogate models and differentiable programming for High Energy Physics detector simulation", Snowmass 2021, arXiv:2203.08806.
- P. Harris et al., "Physics Community Needs, Tools, and Resources for Machine Learning", Snowmass 2021, arXiv:2203.16255.
- A. Ćiprijanović et al., "DeepAstroUDA: Semi-Supervised Universal Domain Adaptation for Cross-Survey Galaxy Morphology Classification and Anomaly Detection", Mach. Learn. Sci. Tech. 4 (2023) 025013, arXiv:2302.02005.
- T. Cai et al., "Accelerating Machine Learning Inference with GPUs in ProtoDUNE Data Processing", Comp. Soft. Big Sci. 7 (2023) 11, arXiv:2301.04633.
- O. Amram and K. Pedro, "Denoising diffusion models with geometry adaptation for high fidelity calorimeter simulation", Phys. Rev. D 108 (2023) 072014, arXiv:2308.03876.
- C. Savard et al., "Optimizing High Throughput Inference on Graph Neural Networks at Shared Computing Facilities with the NVIDIA Triton Inference Server", arXiv:2312.06838, submitted to Comp. Soft. Big Sci.
- CMS Collaboration, "Portable acceleration of CMS computing workflows with coprocessors as a service", arXiv:2402.15366, submitted to Comp. Soft. Big Sci.
- S. Bein et al., "Refining fast simulation using machine learning", Eur. Phys. J. Web Conf. 295 (2024) 09032, arXiv:2309.12919.

Dark QCD References

- 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.
- D. Clowe et al., “A direct empirical proof of the existence of dark matter”, Astrophys. J. 648 (2006) L109, arXiv:astro-ph/0608407.
- T. Han, Z. Si, K. M. Zurek, and M. J. Strassler, “Phenomenology of hidden valleys at hadron colliders”, JHEP 07 (2008) 008, arXiv:0712.2041.
- M. J. Strassler, “Why Unparticle Models with Mass Gaps are Examples of Hidden Valleys”, arXiv:0801.0629.
- M. J. Strassler, “On the Phenomenology of Hidden Valleys with Heavy Flavor”, arXiv:0806.2385.
- K. Petraki and R. R. Volkas, “Review of asymmetric dark matter”, Int. J. Mod. Phys. A 28 (2013) 1330028, 1305.4939.
- Y. Bai and P. Schwaller, “Scale of dark QCD”, Phys. Rev. D 89 (2014) 063522, arXiv:1306.4676.
- T. Cohen, M. Lisanti, and H. K. Lou, “Semi-visible jets: Dark matter undercover at the LHC”, Phys. Rev. Lett. 115 (2015) 171804, arXiv:1503.00009.
- P. Schwaller, D. Stolarski, and A. Weiler, “Emerging jets”, JHEP 05 (2015) 059, arXiv:1502.05409.
- G. D. Kribs and E. T. Neil, “Review of strongly-coupled composite dark matter models and lattice simulations”, Int. J. Mod. Phys. A 31 (2016) 1643004, arXiv:1604.04627.
- S. Knapen, S. Pagan Griso, M. Papucci, and D. J. Robinson, “Triggering Soft Bombs at the LHC”, JHEP 08 (2017) 076, arXiv:1612.00850.
- T. Cohen, M. Lisanti, H. K. Lou, and S. Mishra-Sharma, “LHC searches for dark sector showers”, JHEP 11 (2017) 196, arXiv:1707.05326.
- S. Renner and P. Schwaller, “A flavoured dark sector”, JHEP 08 (2018) 052, arXiv:1803.08080.
- N. Aghanim et al., “Planck 2018 results. VI. Cosmological parameters”, Astron. Astrophys. 641 (2020) A6, 1807.06209.
- H. Beauchesne, E. Bertuzzo, and G. Grilli Di Cortona, “Dark matter in Hidden Valley models with stable and unstable light dark mesons”, JHEP 04 (2019) 118, arXiv:1809.10152.
- CMS Collaboration, “Search for new particles decaying to a jet and an emerging jet”, JHEP 02 (2019) 179, arXiv:1810.10069.
- H. Qu and L. Gouskos, “ParticleNet: Jet Tagging via Particle Clouds”, Phys. Rev. D 101 (2020) 056019, arXiv:1902.08570.
- H. Beauchesne and G. Grilli Di Cortona, “Classification of dark pion multiplets as dark matter candidates and collider phenomenology", JHEP 02 (2020) 196, arXiv:1910.10724.
- CMS Collaboration, “Search for high mass dijet resonances with a new background prediction method in proton-proton collisions at $\sqrt{s}$ = 13 TeV”, JHEP 05 (2020) 033, arXiv:1911.03947.
- T. Cohen, J. Doss, and M. Freytsis, “Jet substructure from dark sector showers”, JHEP 09 (2020) 118, arXiv:2004.00631.
- 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.
- C. Cesarotti, M. Reece, and M. Strassler, “Spheres To Jets: Tuning Event Shapes with 5d Simplified Models”, JHEP 05 (2021) 096, arXiv:2009.08981.
- S. Knapen, J. Shelton, and D. Xu, “Perturbative benchmark models for a dark shower search program”, Phys. Rev. D 103 (2021) 115013, arXiv:2103.01238.
- D. Kim, K. Kong, K. T. Matchev, M. Park and P. Shyamsundar, “Deep-learned event variables for collider phenomenology”, Phys. Rev. D 107 (2023) L031904, arXiv:2105.10126.
- CMS Collaboration, “Search for new particles in events with energetic jets and large missing transverse momentum…”, JHEP 11 (2021) 153, arXiv:2107.13021.
- CMS Collaboration, “Level-1 Trigger Calorimeter Image Convolutional Anomaly Detection Algorithm", CMS-DP-2023-086.
- CMS Collaboration, “Anomaly Detection in the CMS Global Trigger Test Crate for Run 3", CMS-DP-2023-079.
- CMS Collaboration, “Search for long-lived particles decaying in the CMS muon detectors in proton-proton collisions at $\sqrt{s}$ = 13 TeV”, arXiv:2402.01898, submitted to Phys. Rev. D.