Accelerating Discovery in High Energy Physics using AI

vmikuni@lbl.gov

vinicius-mikuni

Vinicius M. Mikuni

Introduction

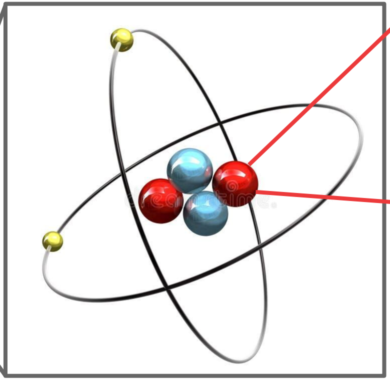
Periodic table of the elements

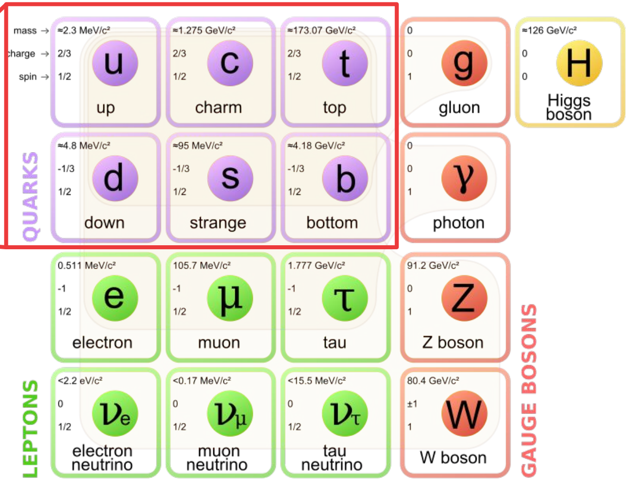
- Alkali metals
- Alkaline-earth metals
- Transition metals
- Other metals
- Other nonmetals
- Halogens
- Noble gases
- Rare-earth elements (21, 39, 57–71) and lanthanoid elements (57–71 only)
- Actinoid elements

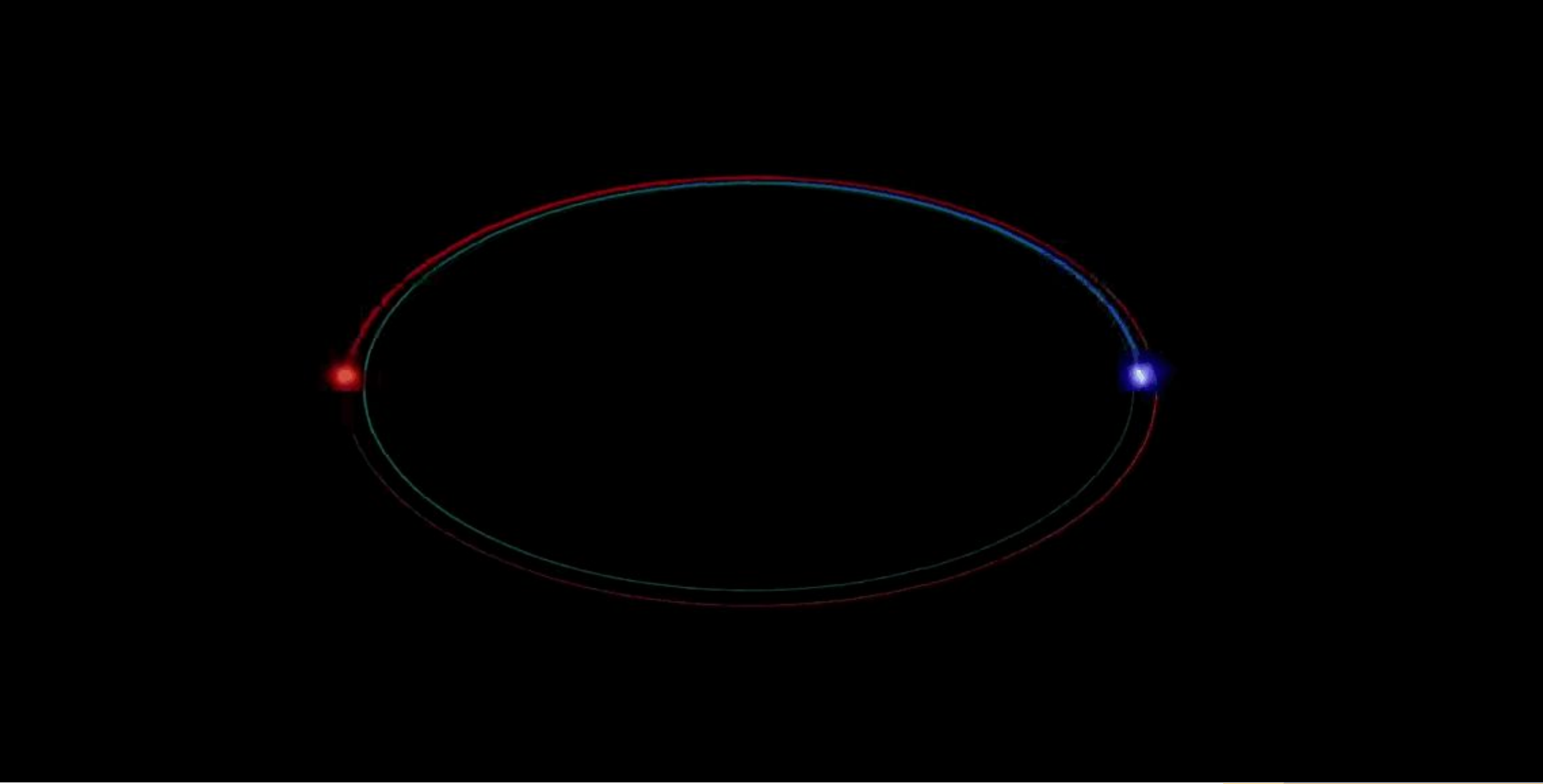
period \ group	1*	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1 H																	2 He
2	3 Li	4 Be											5 B	6 C	7 N	8 O	9 F	10 Ne
3	11 Na	12 Mg											13 Al	14 Si	15 P	16 S	17 Cl	18 Ar
4	19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe	27 Co	28 Ni	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
5	37 Rb	38 Sr	39 Y	40 Zr	41 Nb	42 Mo	43 Tc	44 Ru	45 Rh	46 Pd	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
6	55 Cs	56 Ba	57 La	72 Hf	73 Ta	74 W	75 Re	76 Os	77 Ir	78 Pt	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85 At	86 Rn
7	87 Fr	88 Ra	89 Ac	104 Rf	105 Db	106 Sg	107 Bh	108 Hs	109 Mt	110 Ds	111 Rg	112 Cn	113 Nh	114 Fl	115 Mc	116 Lv	117 Ts	118 Og

lanthanoid series 6	58 Ce	59 Pr	60 Nd	61 Pm	62 Sm	63 Eu	64 Gd	65 Tb	66 Dy	67 Ho	68 Er	69 Tm	70 Yb	71 Lu
actinoid series 7	90 Th	91 Pa	92 U	93 Np	94 Pu	95 Am	96 Cm	97 Bk	98 Cf	99 Es	100 Fm	101 Md	102 No	103 Lr

*Numbering system adopted by the International Union of Pure and Applied Chemistry (IUPAC).



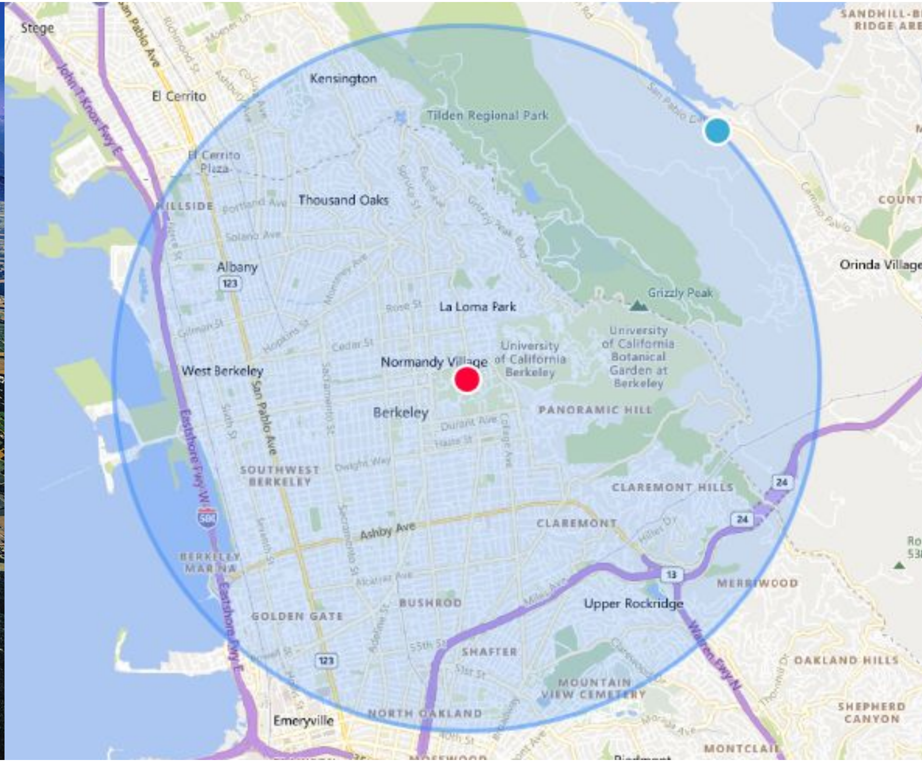




Particle colliders

Animation from [business insider]

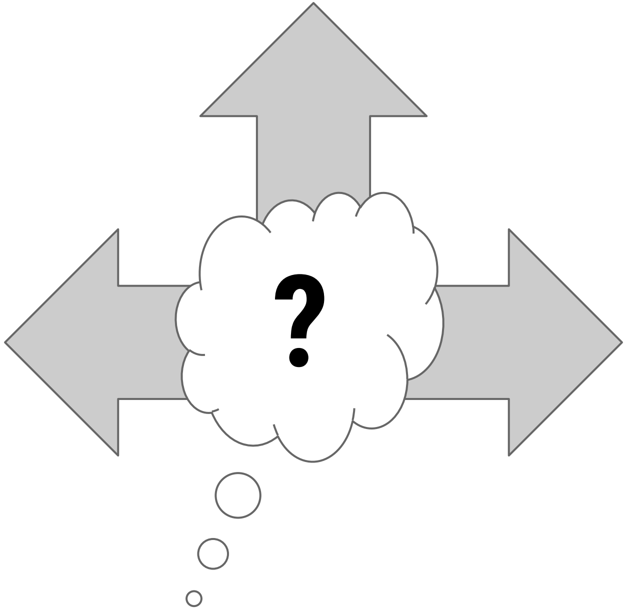
https://www.calcmaps.com/map-radius/



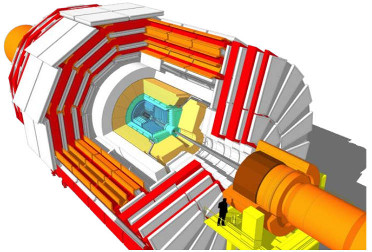
The Large Hadron Collider (**LHC**) is a 3 mile radius accelerator facility, accelerating particles near the speed of light

Discovery Challenge

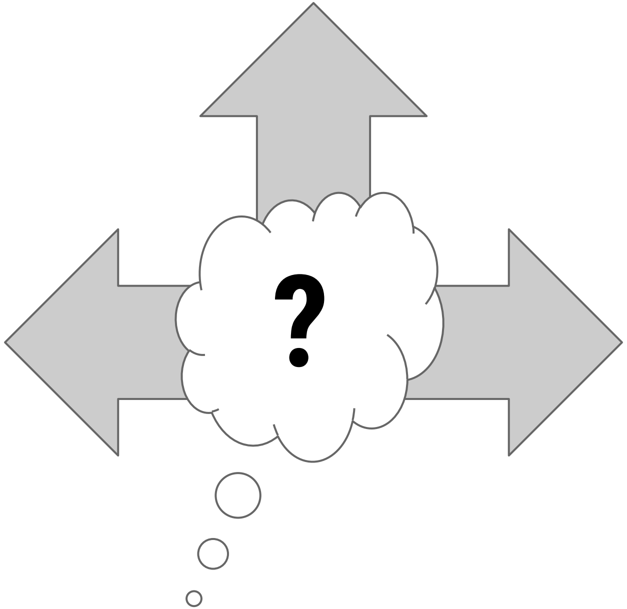
Computing Challenge



Measurement Challenge



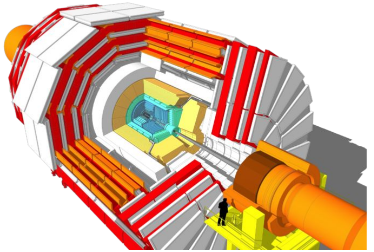
Discovery Challenge



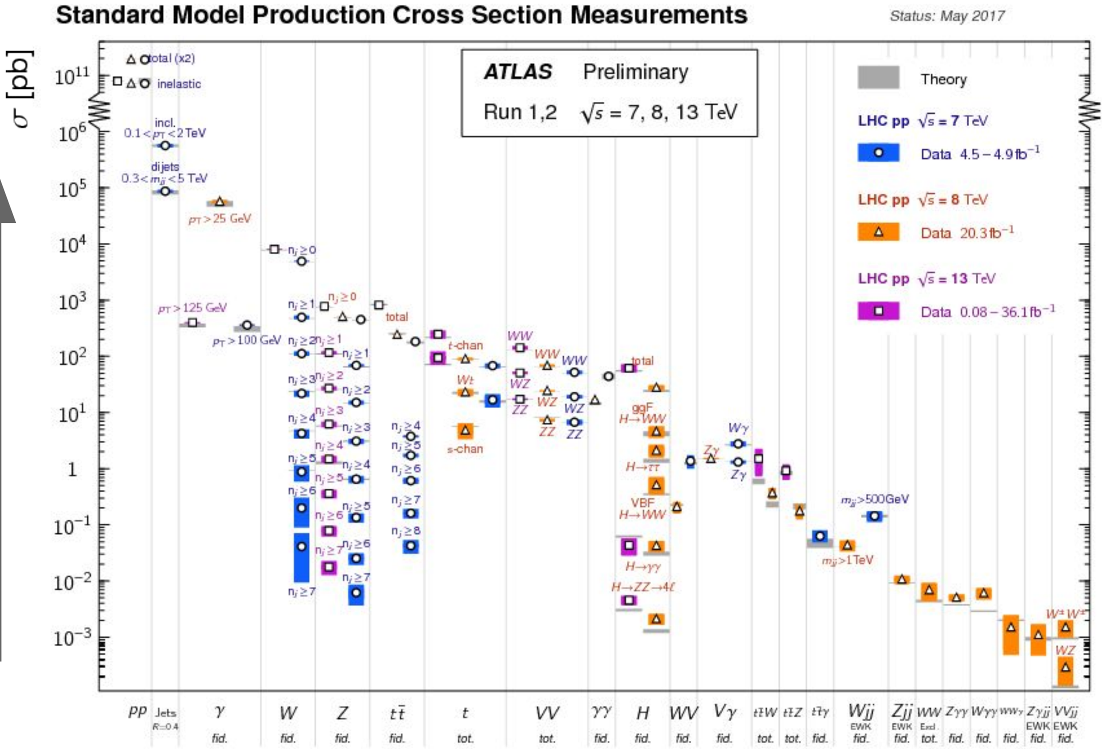
Computing Challenge

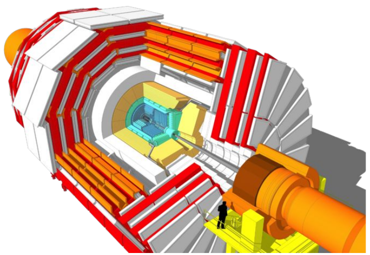
Measurement Challenge

The Challenge

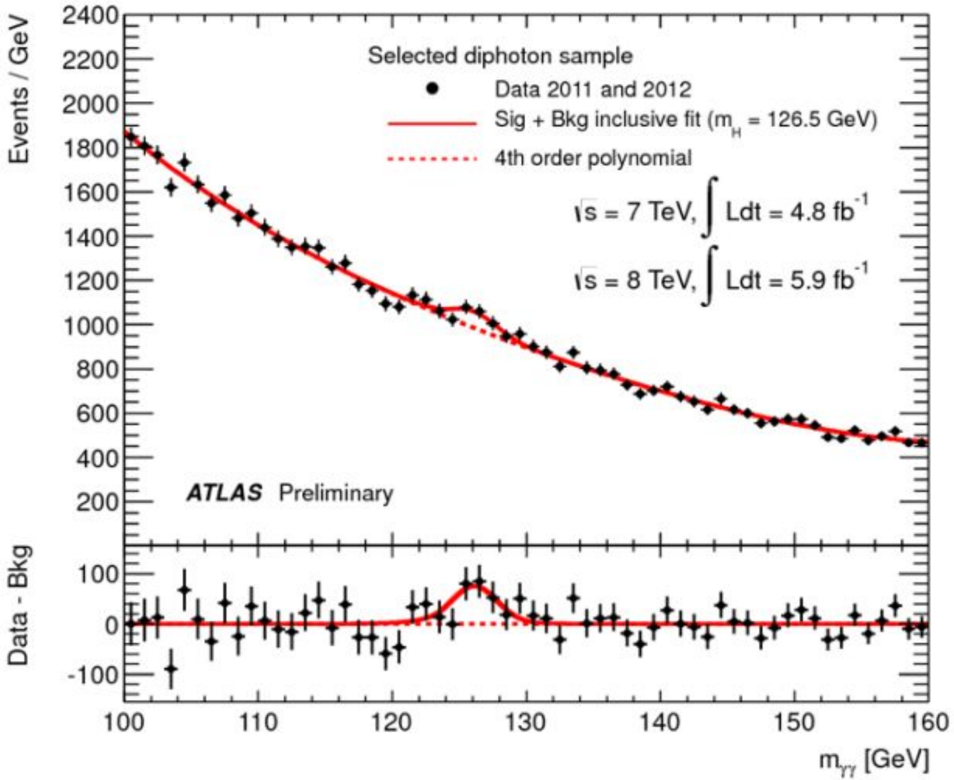


More Likely to Happen





The Challenge



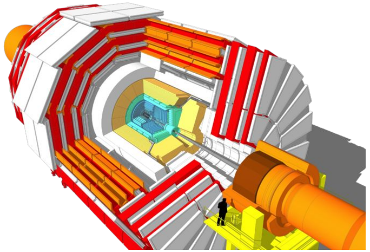
Only about **1 in 10 billion** collisions at the LHC produce a Higgs Boson

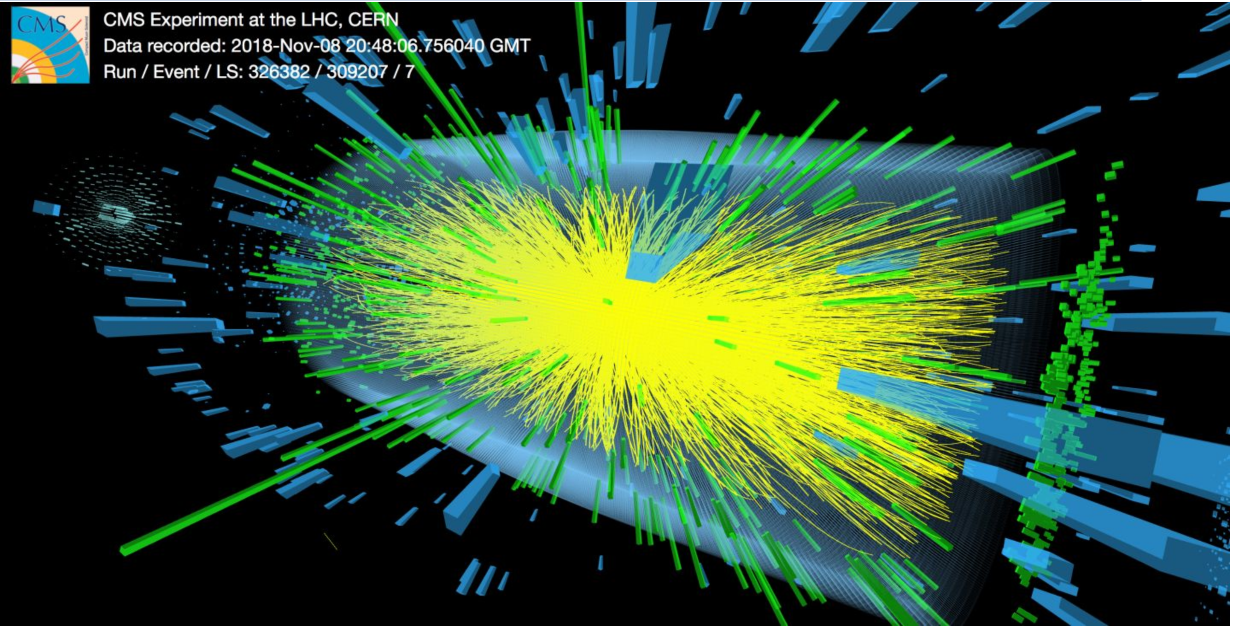
Comparison:

- Odds of being struck by lightning: **1 in 15 thousand**
- Odds of being killed by a vending machine: **1 in 112 million**
- Odds of winning the Powerball: **1 in 300 million**

Source: https://stacker.com/art-culture/odds-50-random-events-happening-you

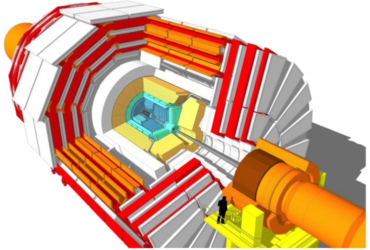
The Challenge

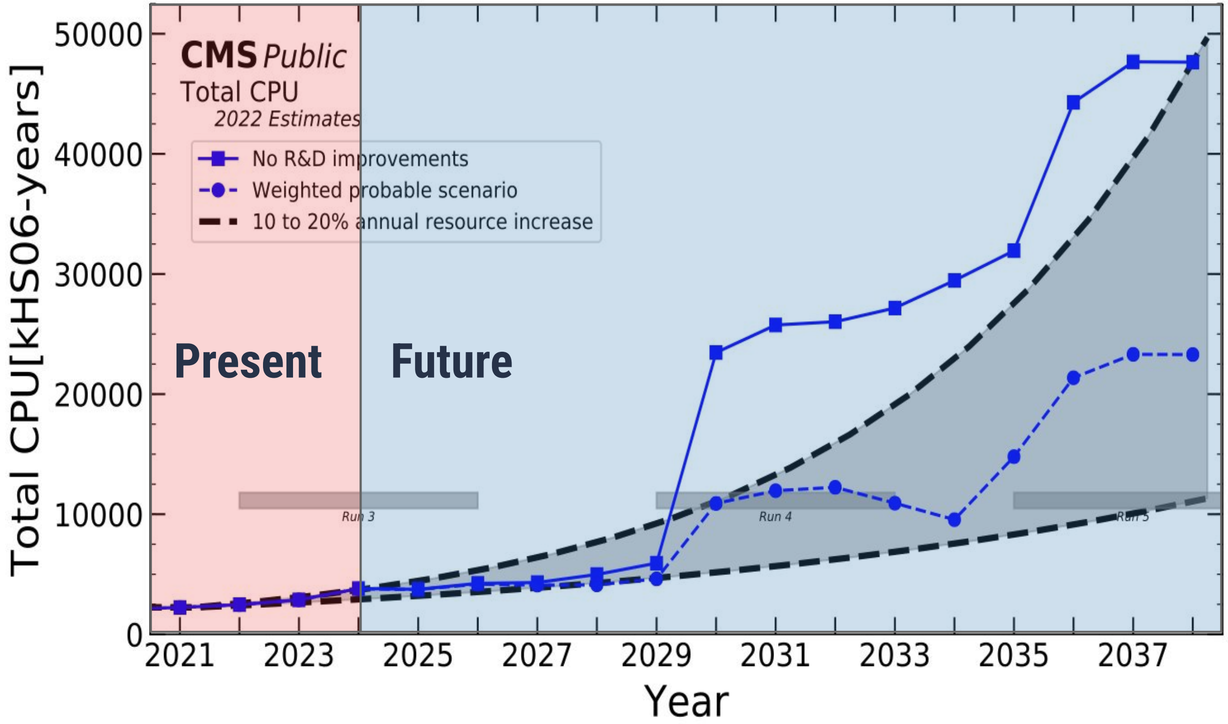




The trick: Bunches of protons cross each other every **25 ns**, resulting in about **600 Million effective collisions per second**

The Challenge

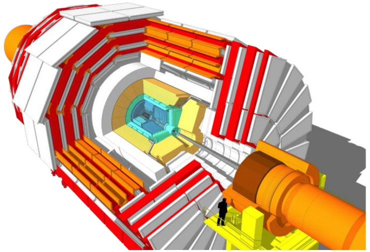




Future upgrades of the LHC experiment will aim to increase the likelihood of collisions happening, **exceeding the current computing budget**

Source: CMS-NOTE-2022-008

The Challenge

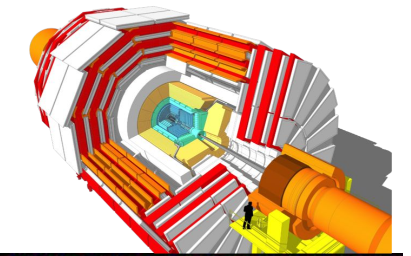


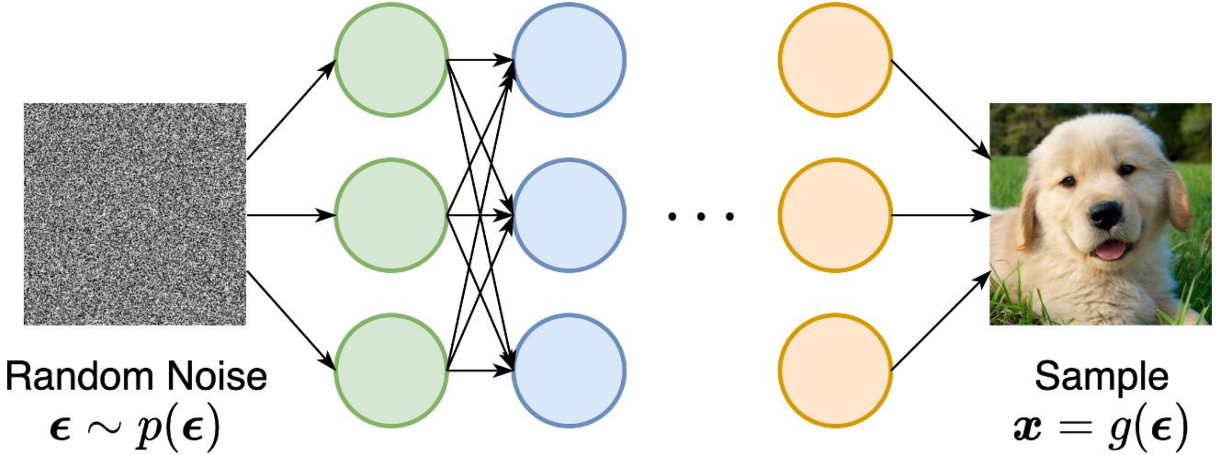
Physics Theory → Physics Simulation → Detector Simulator → Simulated data

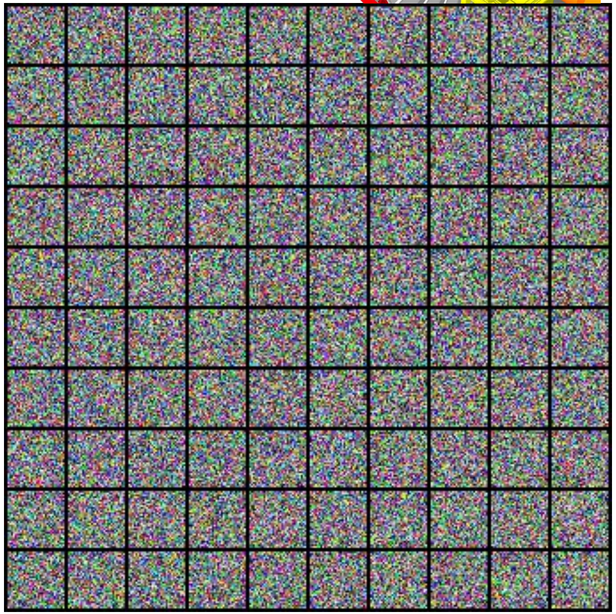
Nature → Detector interaction → Measured data

We can only compare our physics predictions with experiments through the use of **simulations**. Detector simulation takes more than **40%** of the computing resources

Generative models



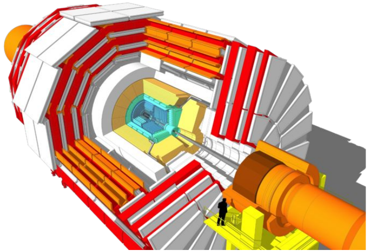


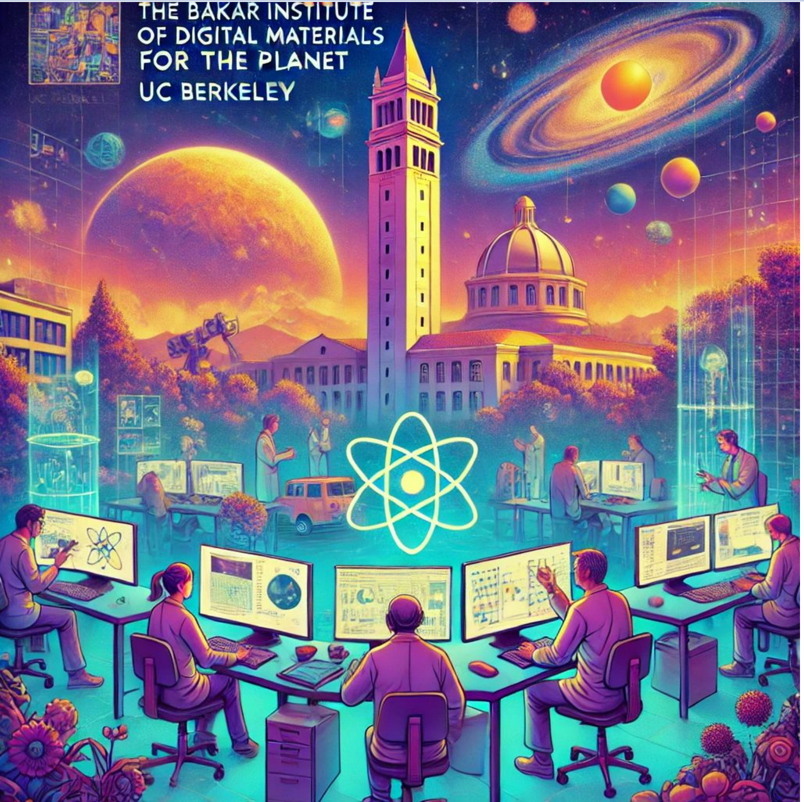


Source: https://yang-song.net/blog/2021/score/

Generative models are a class of algorithms trained to transform easy-to-sample noise into data

Diffusion Generative Models

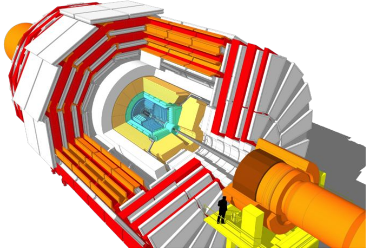




"Scientists from BIDMAP working on Science and Machine Learning"

https://openai.com/dall-e-2

Diffusion Generative Models



Forward SDE (data → noise)

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

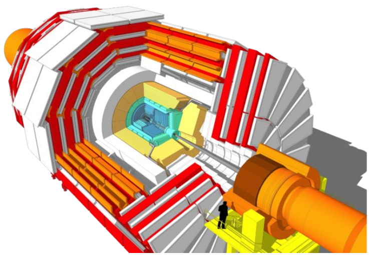
score function

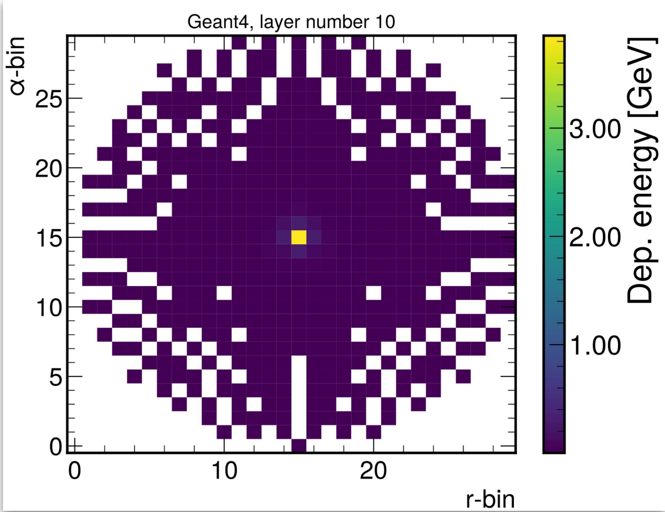
$$\mathbf{x}(0) \longleftarrow \mathrm{d}\mathbf{x} = \left[\mathbf{f}(\mathbf{x}, t) - g^2(t)\boxed{\nabla_{\mathbf{x}} \log p_t(\mathbf{x})}\right] \mathrm{d}t + g(t)\mathrm{d}\bar{\mathbf{w}} \longleftarrow \mathbf{x}(T)$$

Reverse SDE (noise → data)

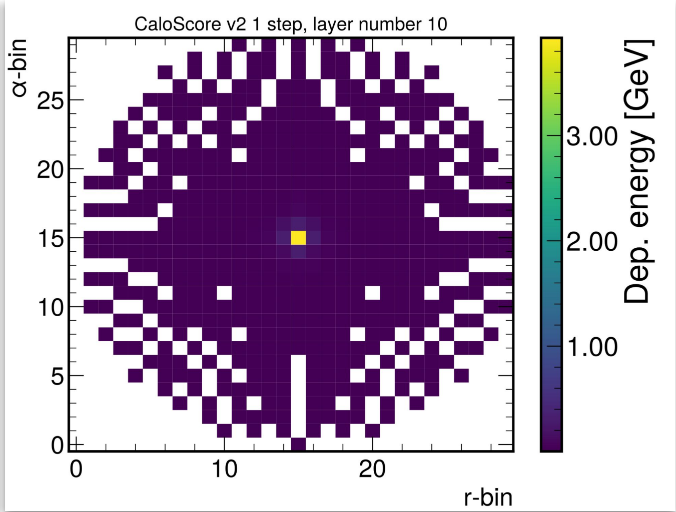
Source: https://yang-song.net/blog/2021/score/

Diffusion Generative Models for Detector Simulation





Physics Simulation

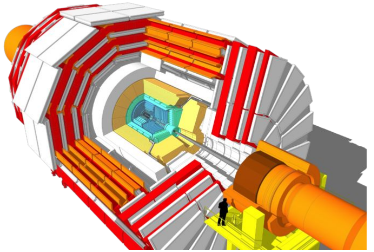


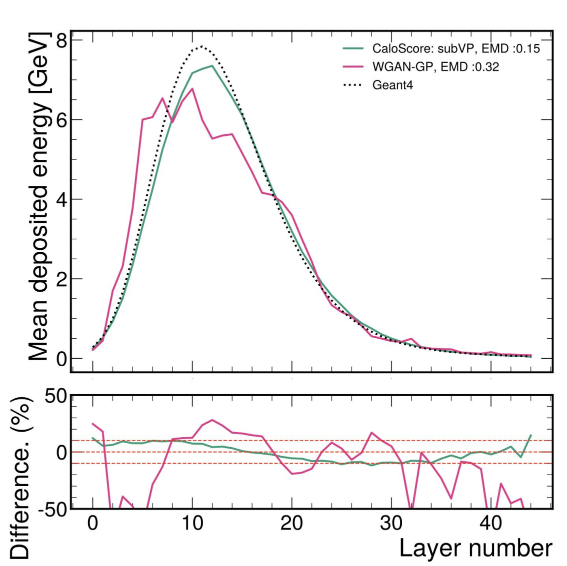
Generated by **CaloScore**

First Diffusion model in High Energy Physics named **CaloScore**.
Up to 50k Detector Components simulated

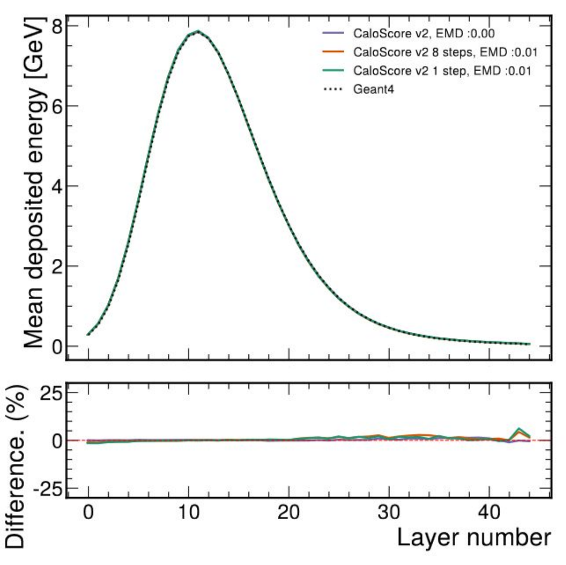
- **V. Mikuni** and B. Nachman Phys. Rev. D **106**, 092009
- **V. Mikuni** and B. Nachman 2024 *JINST* **19** P02001

Diffusion Generative Models for Detector Simulation





Energy deposition inferred from sum of pixels

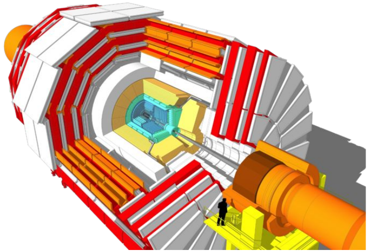


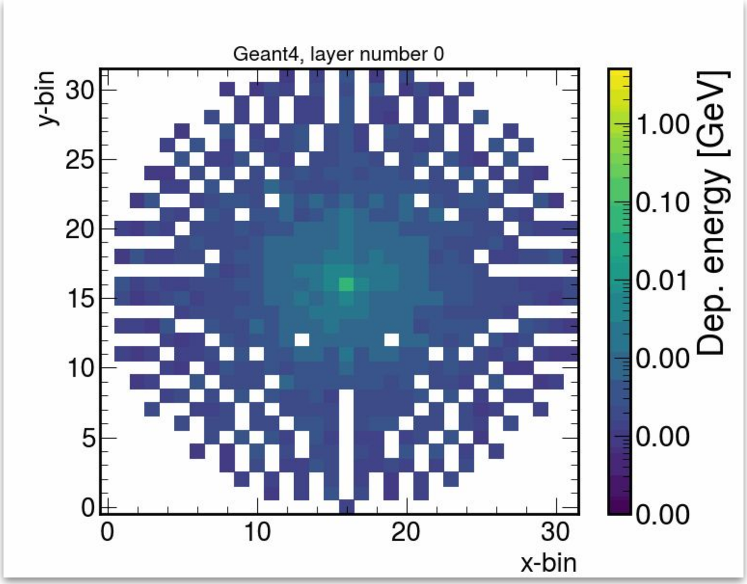
Additional model trained to learn the energy sum

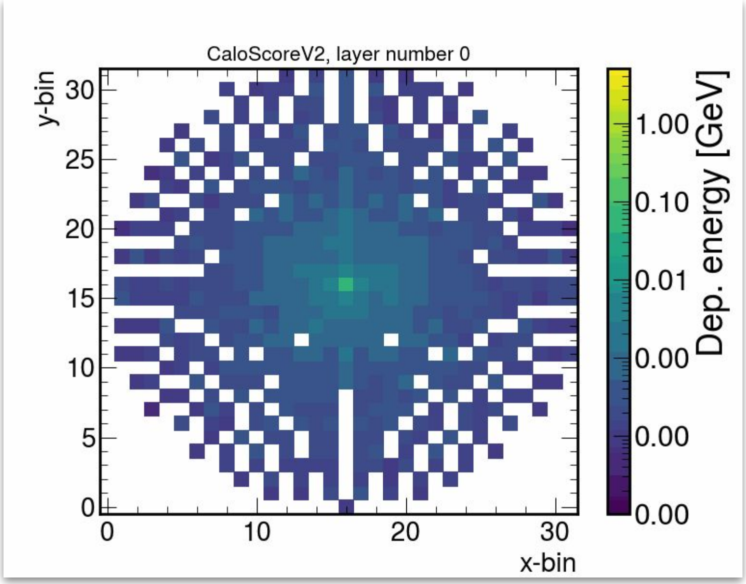
Improve energy conservation by training **2 conditional diffusion models**: One on normalized pixel responses and one to determine the total energy deposition

$$\nabla \log p(x_{\mathrm{norm}}, E) = \nabla \log p(x_{\mathrm{norm}}|E) + \nabla \log p(E)$$

Diffusion Generative Models for Detector Simulation



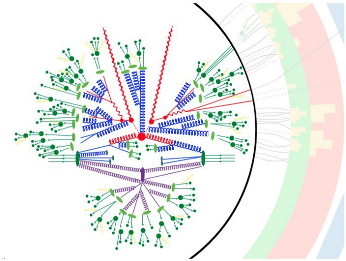




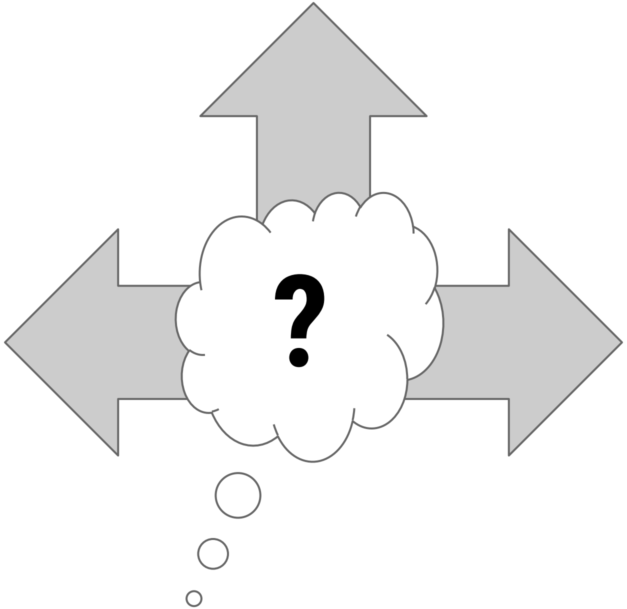
10^5-10^6 times faster than full physics simulation!

Physics Simulator

CaloScore



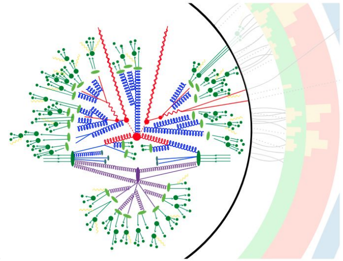
Discovery Challenge



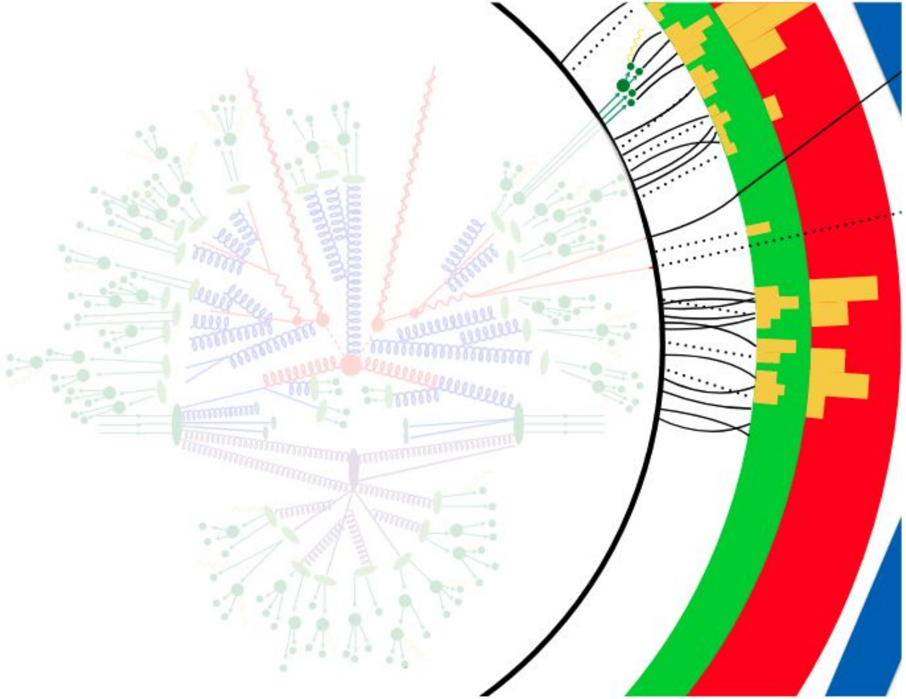
Computing Challenge

Measurement Challenge

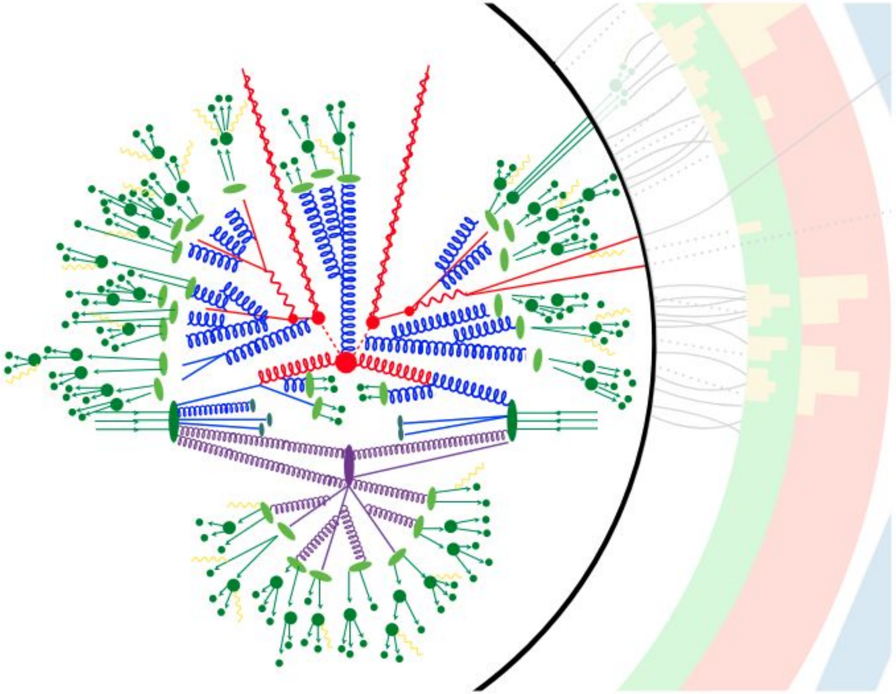
Unfolding



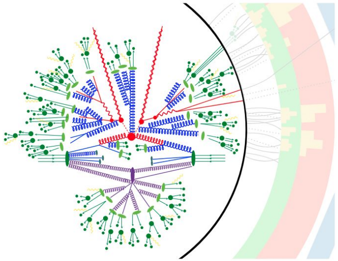
What we measure



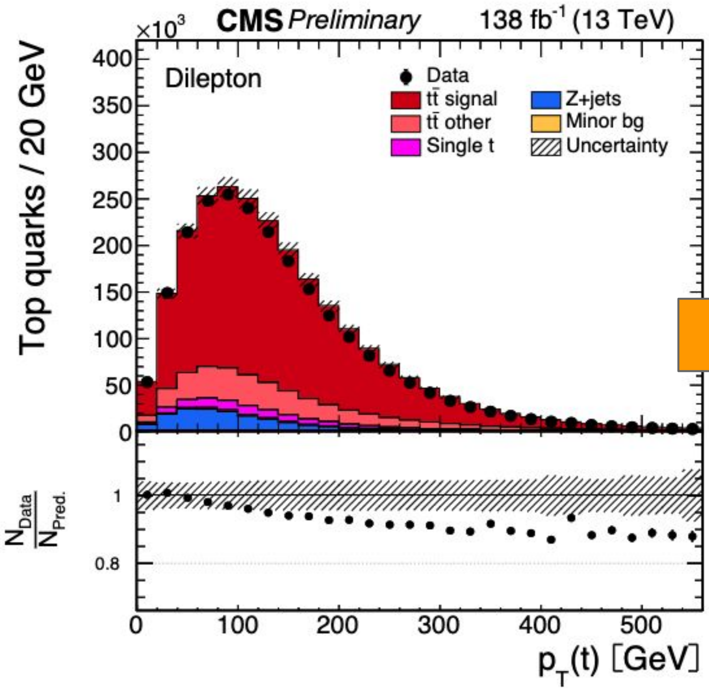
What we want

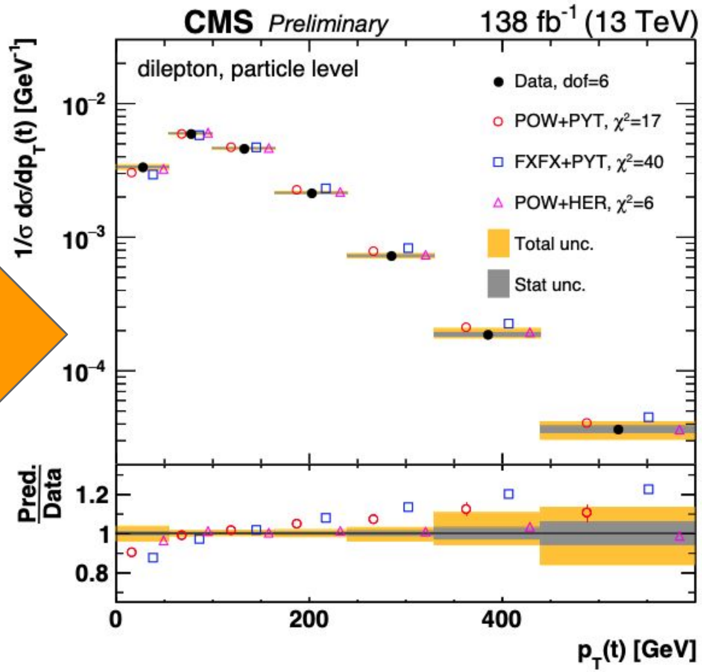


Unfolding



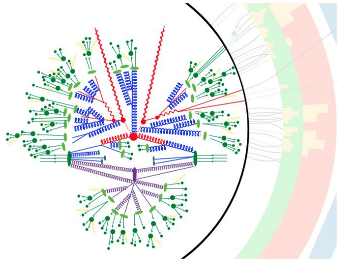
Source: CMS-PAS-TOP-20-006





Traditional methods for **unfolding** use **histograms**

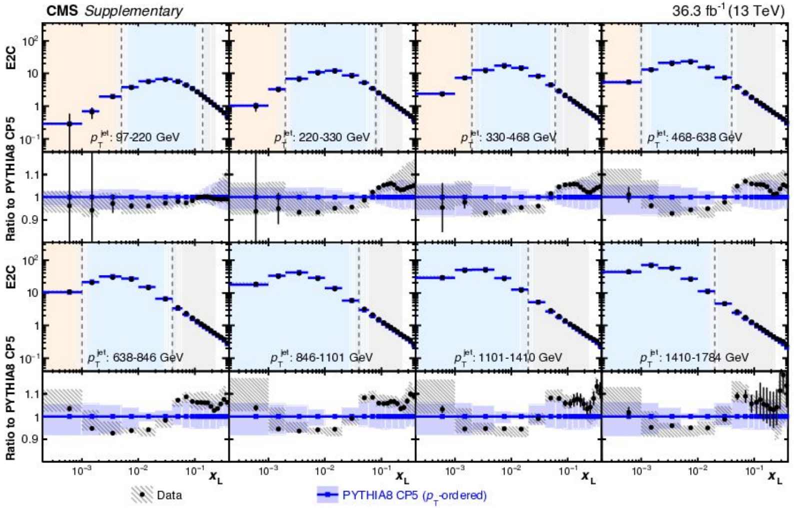
The Challenge



Energy Correlations in Electron-Positron Annihilation: Testing Quantum Chromodynamics

C. Louis Basham, Lowell S. Brown, Stephen D. Ellis, and Sherwin T. Love
Department of Physics, University of Washington, Seattle, Washington 98195
(Received 21 August 1978)

An experimental measure is presented for a precise test of quantum chromodynamics. This measure involves the asymmetry in the energy-weighted opening angles of the jets of hadrons produced in the process $e^+e^- \to$ hadrons at energy W. It is special for several reasons: It is reliably calculable in asymptotically free perturbation theory; it has rapidly vanishing (order $1/W^2$) corrections due to nonperturbative confinement effects; and it is straightforward to determine experimentally.



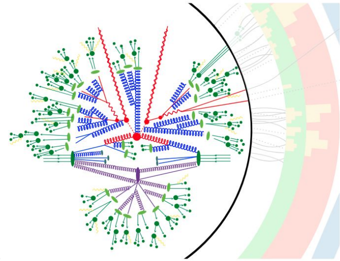
Not everything is naturally represented by a histogram!

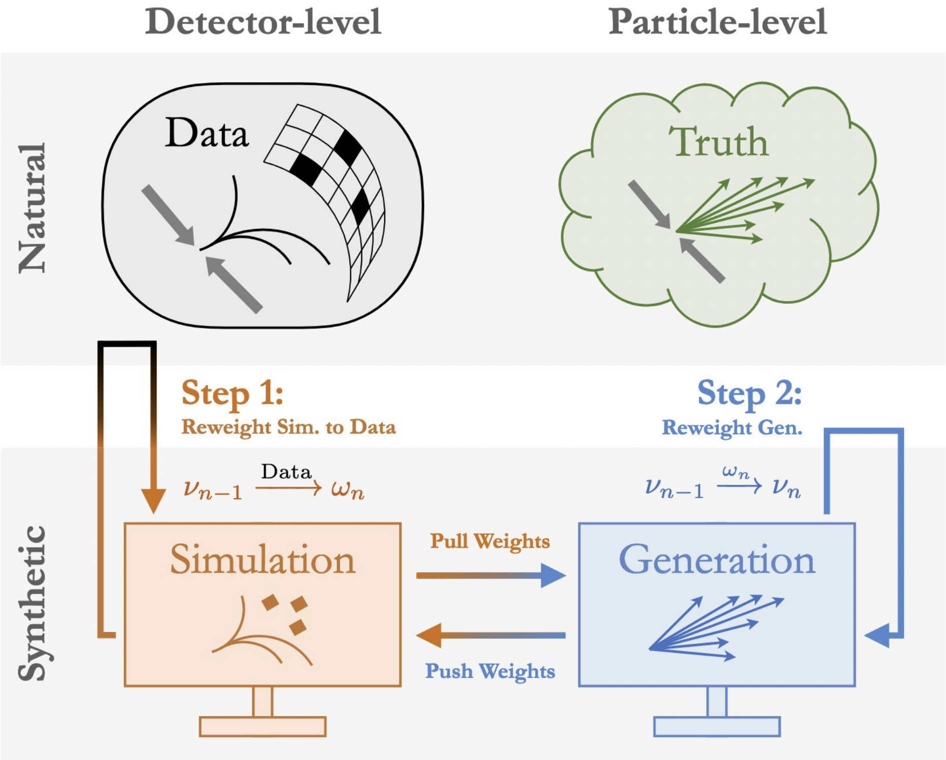
- Some observables are **not** simply counts

2024

CMS-SMP-22-015

OmniFold



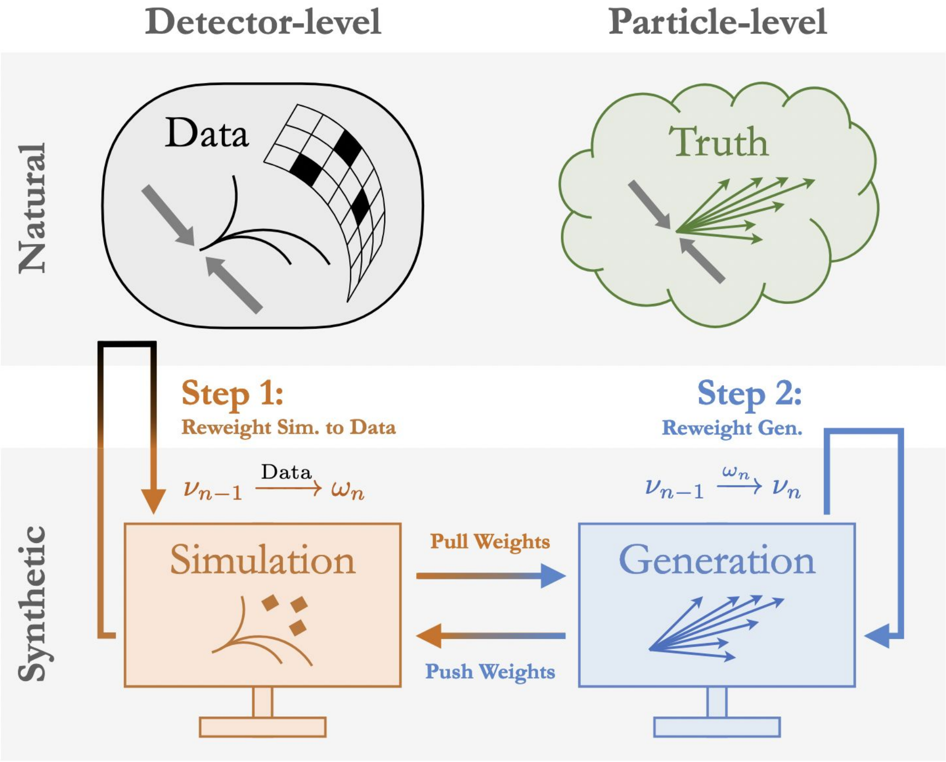


2-step iterative process

- **Step 1**: Reweight simulations to look like data
- **Step 2**: Convert learned weights into functions of particle level objects

Source: Andreassen et al. PRL 124, 182001 (2020)

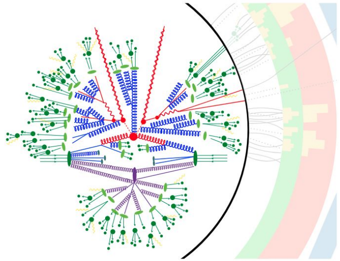
OmniFold



2-step iterative process

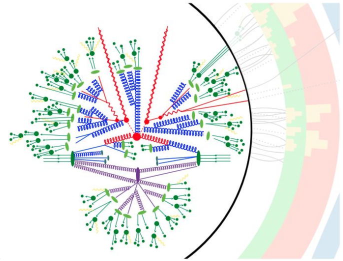
- **Step 1**: Reweight simulations to look like data
- **Step 2**: Convert learned weights into functions of particle level objects
- Use **classifiers** to learn the **reweighting** functions!

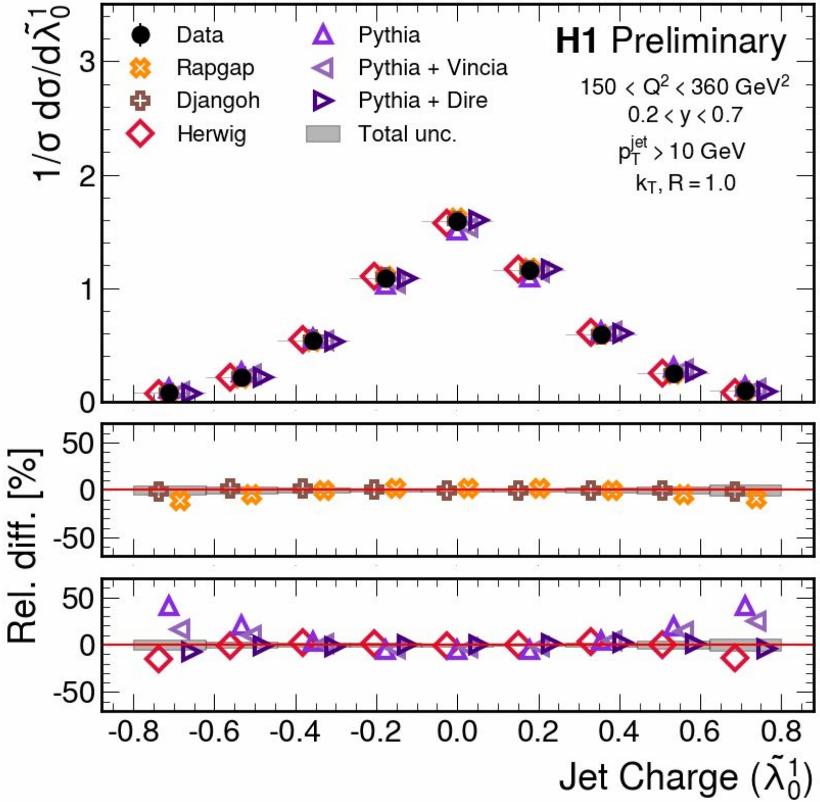
Source: Andreassen et al. PRL 124, 182001 (2020)



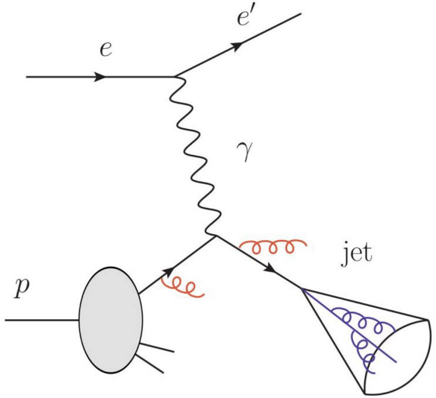
In practice

Histograms are not used during the measurement, only to **display** the results



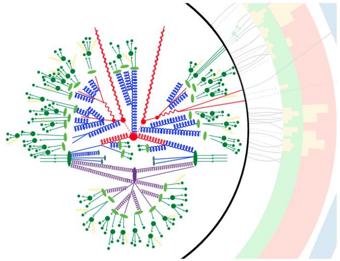


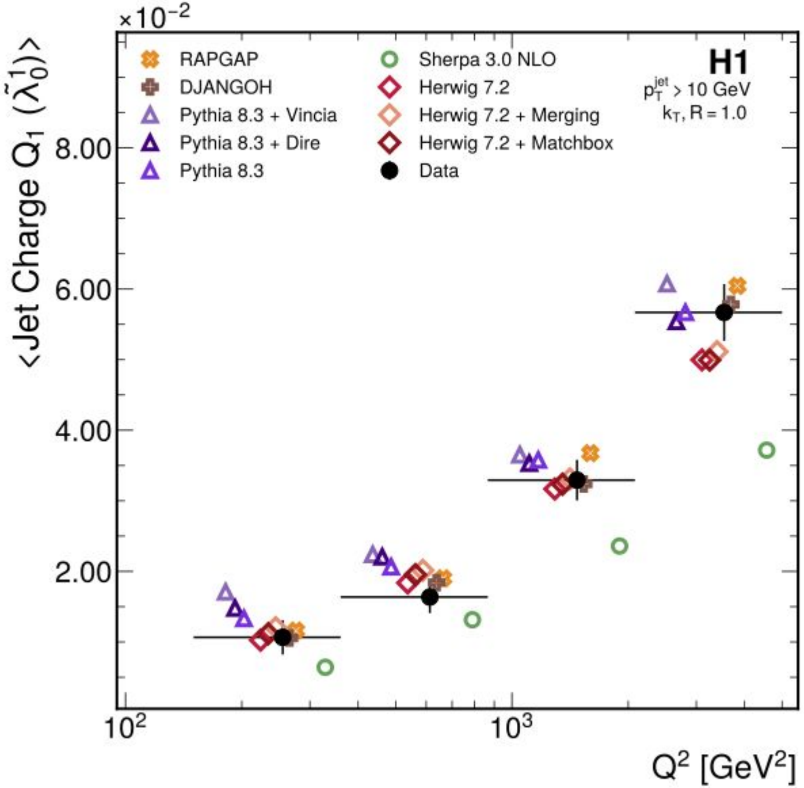
As the energy scale increases, so does the likelihood of scattering a valence quark from the proton, resulting in more **positive jet charges**!



H1 Collaboration: PLB 844 (2023) 138101

In practice

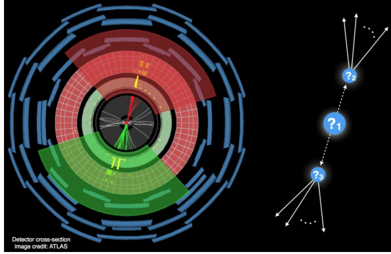




We can quantify this statement by looking at the average jet charge versus energy scale

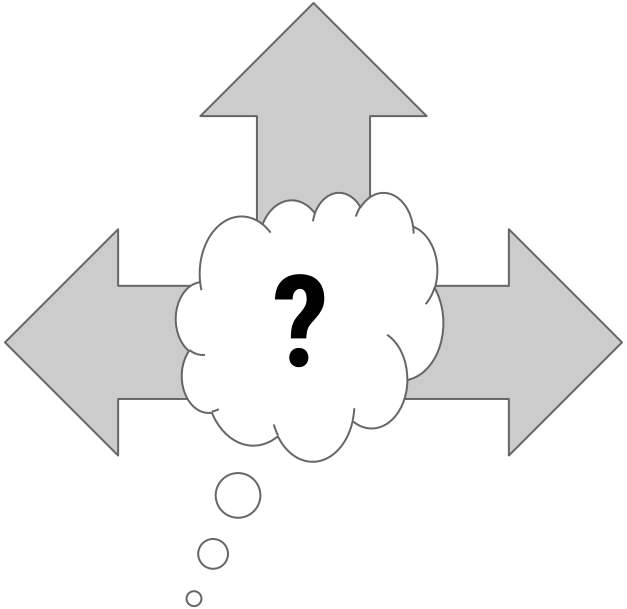
No histograms needed!

Source: Andreassen et al. PRL 124, 182001 (2020)



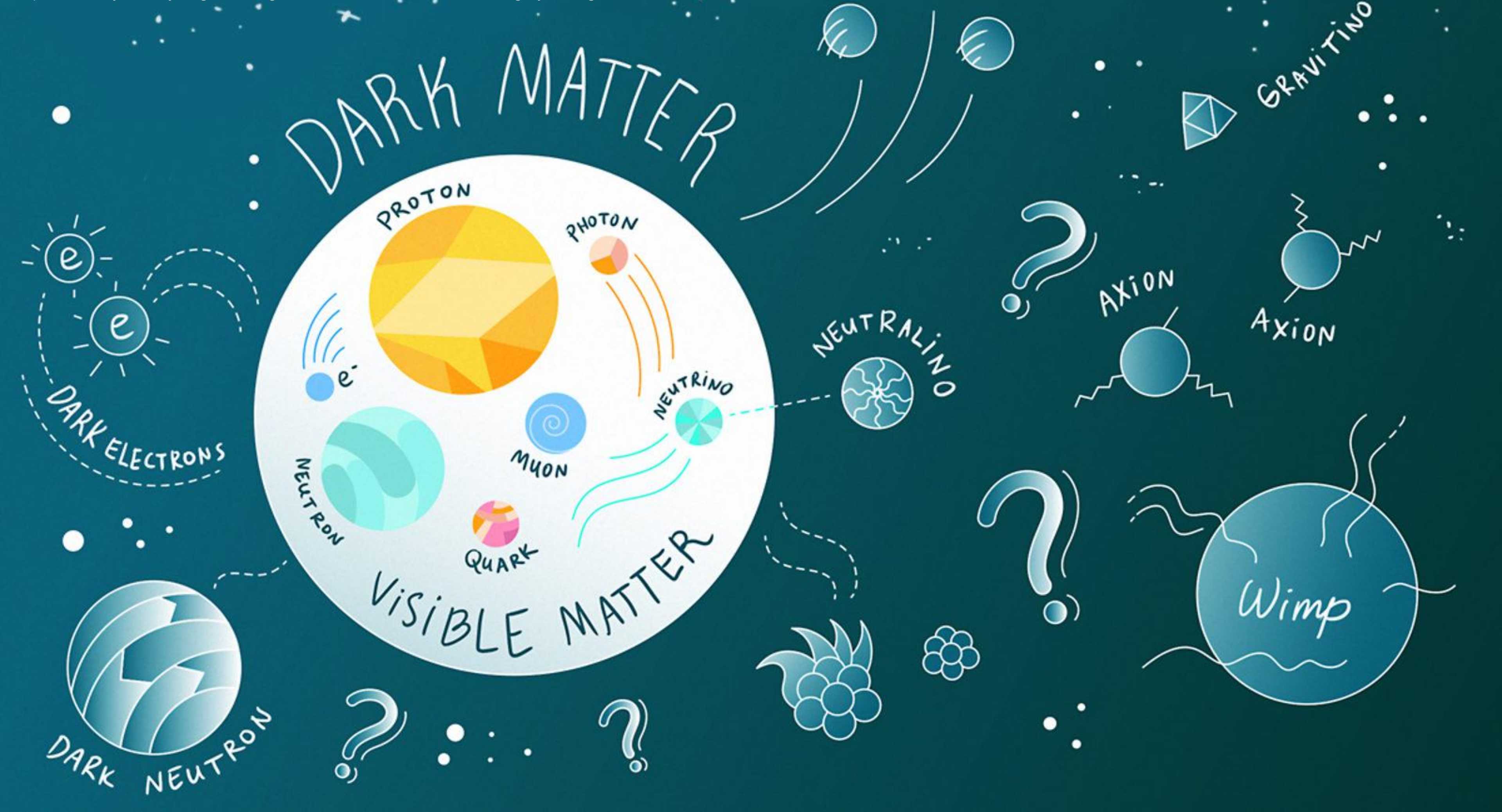
Discovery Challenge

Computing Challenge

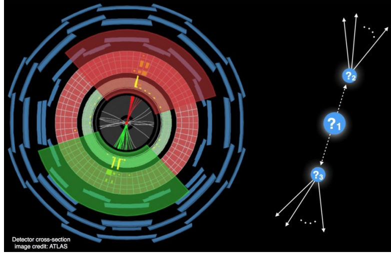


Measurement Challenge

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

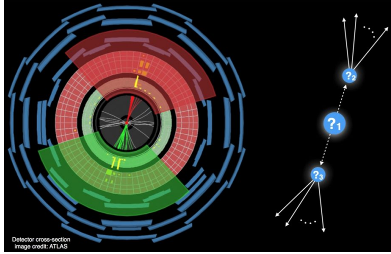


The Challenge



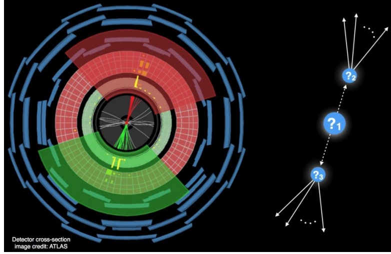
Q: How do you find a needle in a haystack?

The Challenge



Q: How do you find a needle in a haystack?

A: You use a magnet!



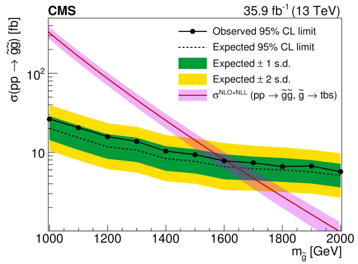
The Challenge

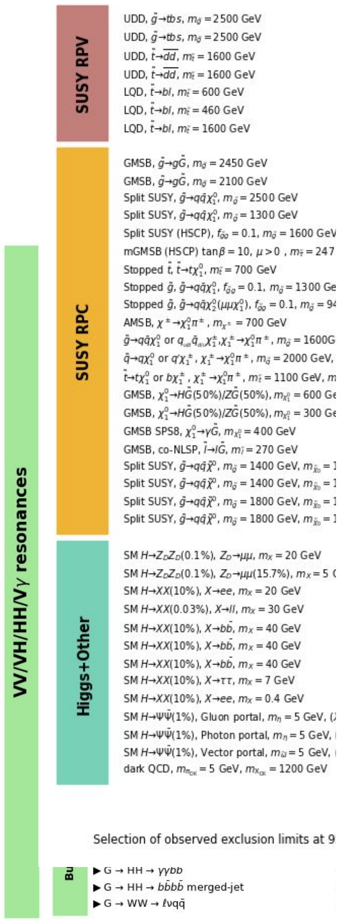
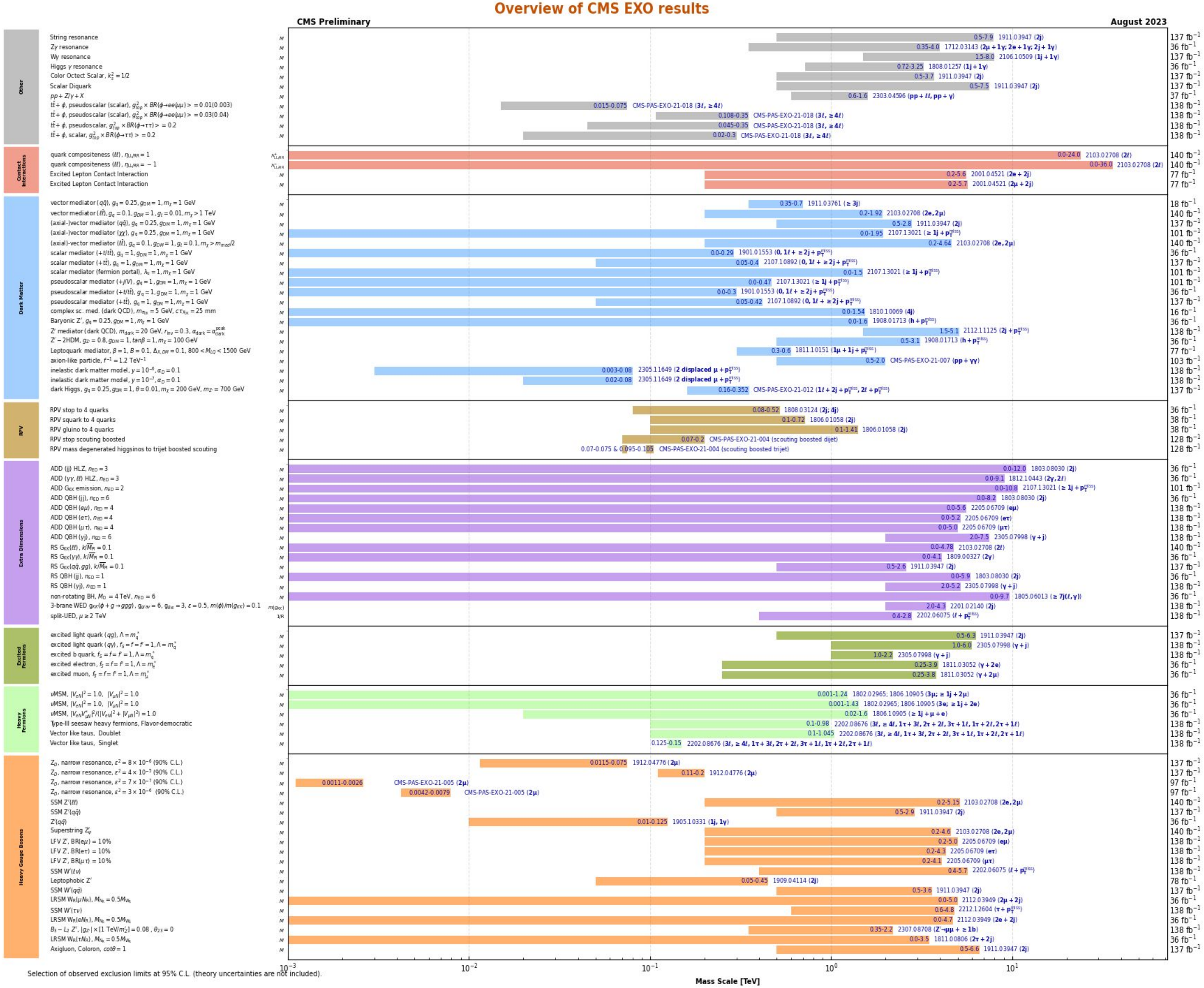
Plausible Theory:
SUSY, WIMPs, LLPs

Verification: Confirm the theory using data

Theory was right!

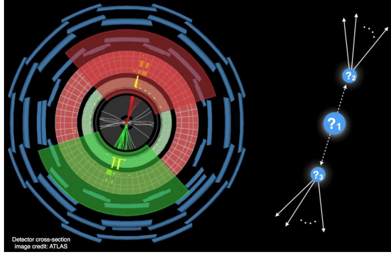
Constrain the new theory




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, $\tilde{g}\to q\bar{q}\tilde{\chi}_1^0$, $m_{\tilde{g}} = 2500$ GeV
Split SUSY, $\tilde{g}\to q\bar{q}\tilde{\chi}_1^0$, $m_{\tilde{g}} = 1300$ 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 $\tilde{t}$, $\tilde{t}\to t\tilde{\chi}_1^0$, $m_{\tilde{t}} = 700$ GeV
Stopped $\tilde{g}$, $\tilde{g}\to q\bar{q}\tilde{\chi}_1^0$, $f_{\tilde{g}g} = 0.1$, $m_{\tilde{g}} = 1300$ Ge
AMSB, $\chi^{\pm}\to\chi_1^0\pi^{\pm}$, $m_{\chi^{\pm}} = 700$ GeV
GMSB SPS8, $\chi_1^0\to\gamma\tilde{G}$, $m_{\chi_1^0} = 400$ GeV
GMSB, co-NLSP, $\tilde{l}\to l\tilde{G}$, $m_{\tilde{l}} = 270$ GeV
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
VV/VH/HH/Vγ resonances
Selection of observed exclusion limits at 9
▶ G → HH → $\gamma\gamma bb$
▶ 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
String resonance 0.5-7.9 1911.03947 (2j) 137 fb^{-1}
Zγ resonance 0.35-4.0 1712.03143 36 fb^{-1}
Wγ resonance 1.5-8.0 2106.10509 (1j + 1γ) 137 fb^{-1}
Higgs γ resonance 0.72-3.25 1808.01257 (1j + 1γ) 36 fb^{-1}
Color Octect Scalar 0.5-3.7 1911.03947 (2j) 137 fb^{-1}
Scalar Diquark 0.5-7.5 1911.03947 (2j) 137 fb^{-1}
pp + Z/γ + X 0.6-1.6 2303.04596 37 fb^{-1}
Contact Interactions
quark compositeness 0.0-24.0 2103.02708 (2ℓ) 140 fb^{-1}
quark compositeness 0.0-36.0 2103.02708 (2ℓ) 140 fb^{-1}
Excited Lepton Contact Interaction 0.2-5.6 2001.04521 (2e + 2j) 77 fb^{-1}
Excited Lepton Contact Interaction 0.2-5.7 2001.04521 (2μ + 2j) 77 fb^{-1}
Dark Matter
RPV
RPV stop to 4 quarks 0.08-0.52 1808.03124 (2j; 4j) 36 fb^{-1}
RPV squark to 4 quarks 0.1-0.72 1806.01058 (2j) 38 fb^{-1}
RPV gluino to 4 quarks 0.1-1.41 1806.01058 (2j) 38 fb^{-1}
RPV stop scouting boosted 0.07-0.2 CMS-PAS-EXO-21-004 (scouting boosted dijet) 128 fb^{-1}
Extra Dimensions
Excited Fermions
excited light quark (qg) 0.5-6.3 1911.03947 (2j) 137 fb^{-1}
excited b quark 1.0-2.2 2305.07998 (γ + j) 138 fb^{-1}
excited electron 0.25-3.9 1811.03052 (γ + 2e) 36 fb^{-1}
excited muon 0.25-3.8 1811.03052 (γ + 2μ) 36 fb^{-1}
Heavy Fermions
Heavy Gauge Bosons
SSM Z'(ℓℓ) 0.2-5.15 2103.02708 (2e, 2μ) 140 fb^{-1}
SSM Z'(qq̄) 0.5-2.9 1911.03947 (2j) 137 fb^{-1}
Z'(qq̄) 0.01-0.125 1905.10331 (1j, 1γ) 36 fb^{-1}
Leptophobic Z' 0.05-0.45 1909.04114 (2j) 78 fb^{-1}
SSM W'(qq̄) 0.5-3.6 1911.03947 (2j) 137 fb^{-1}
Axigluon, Coloron, cotθ = 1 0.5-6.6 1911.03947 (2j) 137 fb^{-1}
Selection of observed exclusion limits at 95% C.L. (theory uncertainties are not included).
Mass Scale [TeV]

(13 TeV)
1000
1200
1400

The Challenge

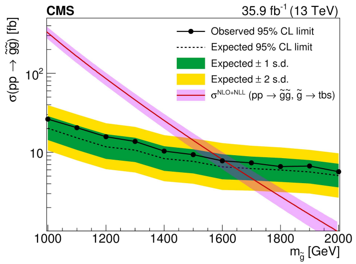


Plausible Theory:
SUSY, WIMPs, LLPs

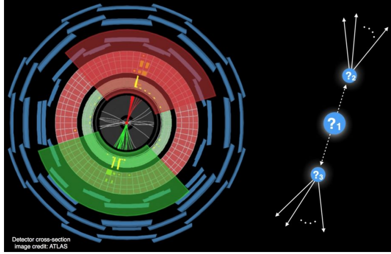
Search for anomalies!

Interpretation

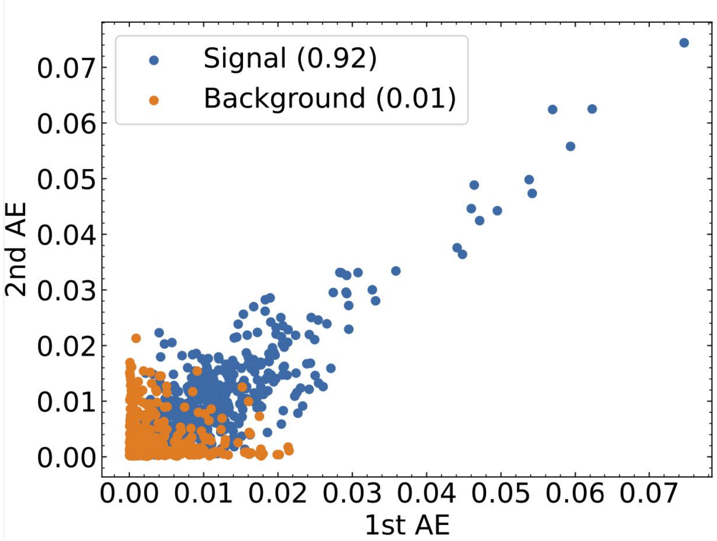
Constrain many theories



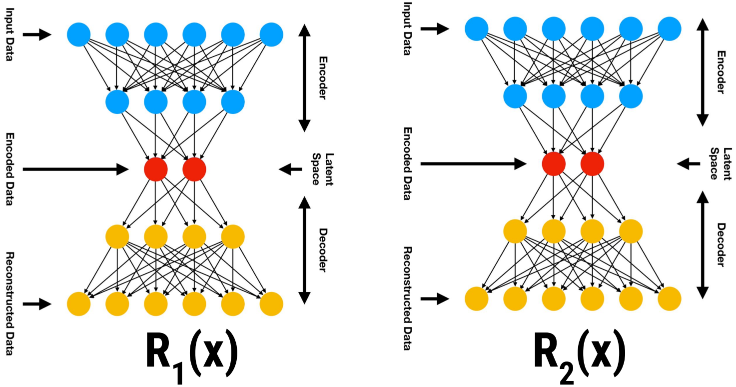
Anomaly detection



- **Autoencoders** learns to **compress** and **decompress** data
- **Anomalies** are often poorly decompressed, yielding a **high reconstruction error**

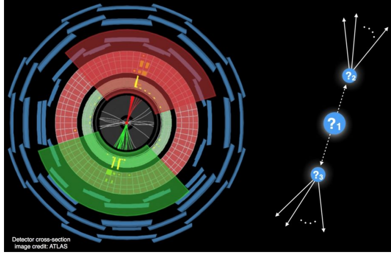


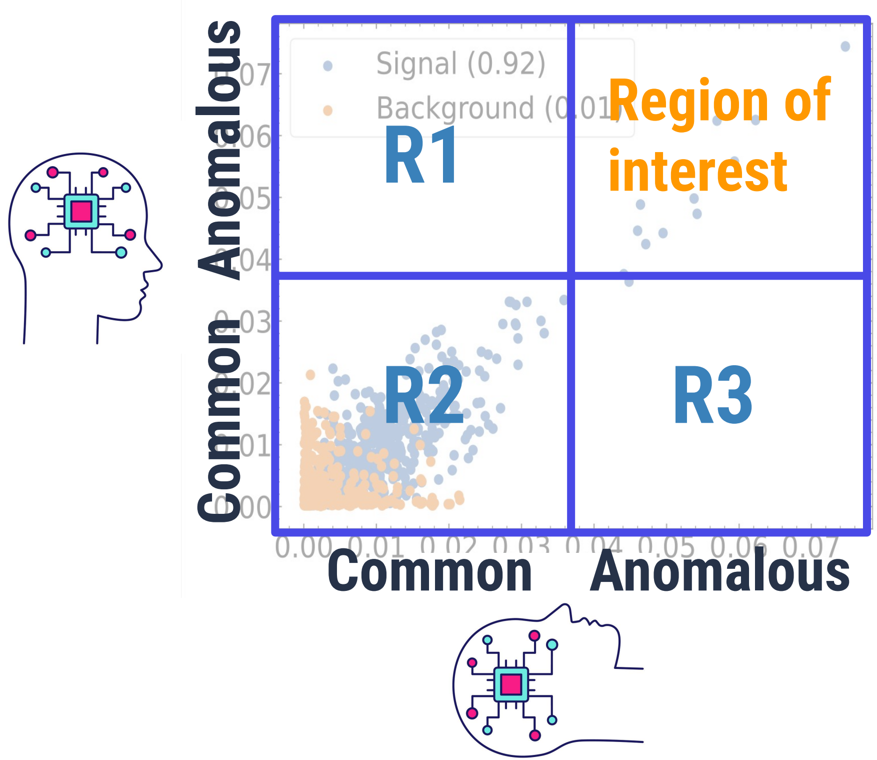
- Train multiple **decorrelated autoencoders**



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

Anomaly detection

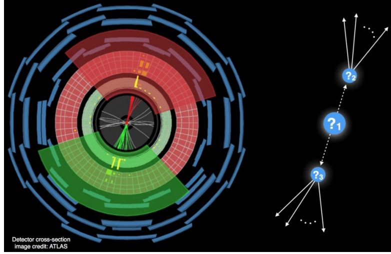




- **Region of interest**: Both autoencoders agree the observation is anomalous
- **Other regions**: Used to estimate the fake rate

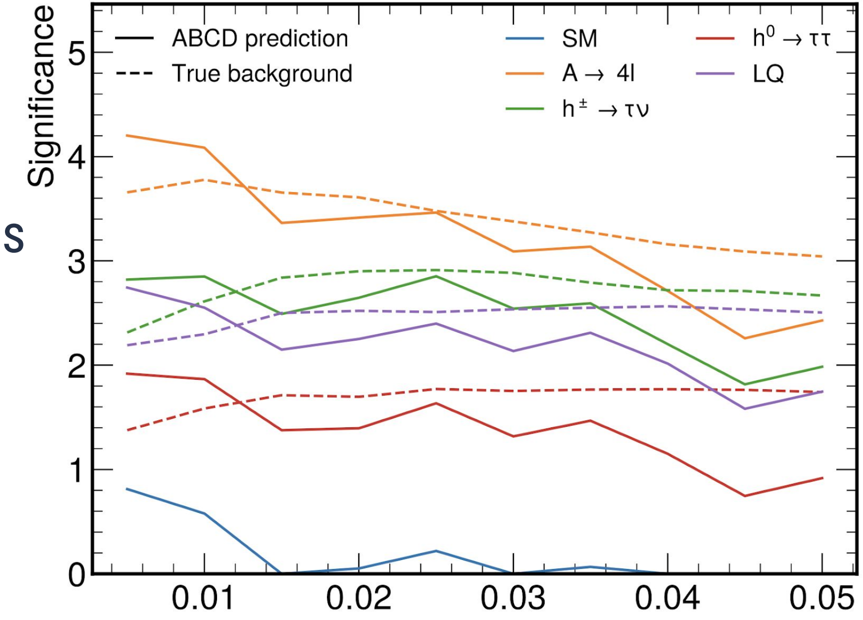
Background in the region of interest = R1*R3/R2

Anomaly detection performance



No anomaly

Other colors: datasets with **0.1%** anomalies and **99.9%** background



In the **absence of new physics**, no excess is observed

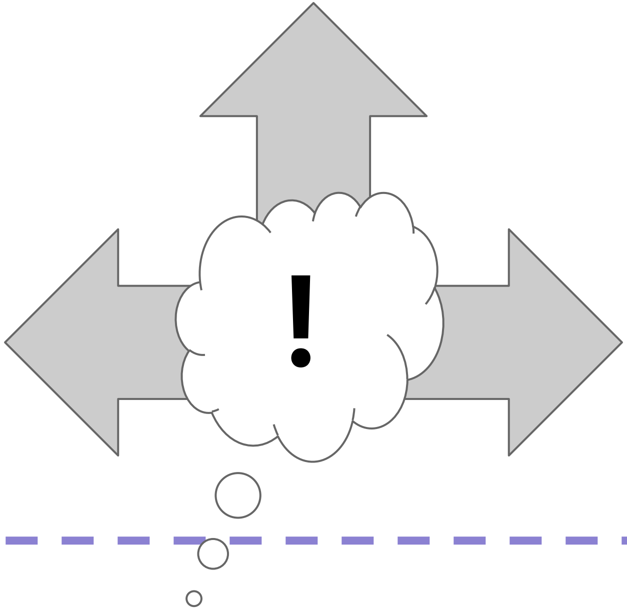
Anomalies identified as an excess translated as a **Significance** or **signal-to-noise ratio**

V. Mikuni, B. Nachman, and D. Shih. *Physical Review D* 105.5 (2022): 055006.

Foundational Models

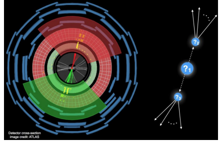
Discovery Challenge

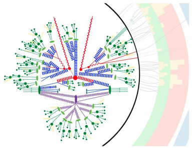
Computing Challenge

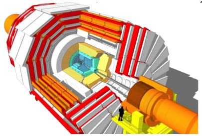


Measurement Challenge

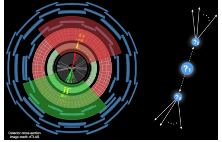
Foundational Models

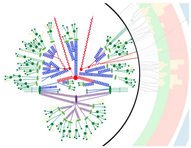


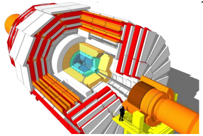




- Foundational models are everywhere now
- In essence, these models are trained on large datasets and can be used for multiple tasks
- How does a **foundational model for science** looks like?





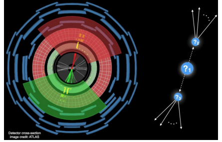
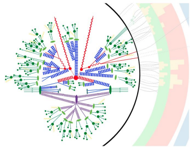
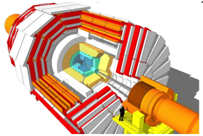


Data

Model

Learning

Present

Data

Lorem ipsum dolor

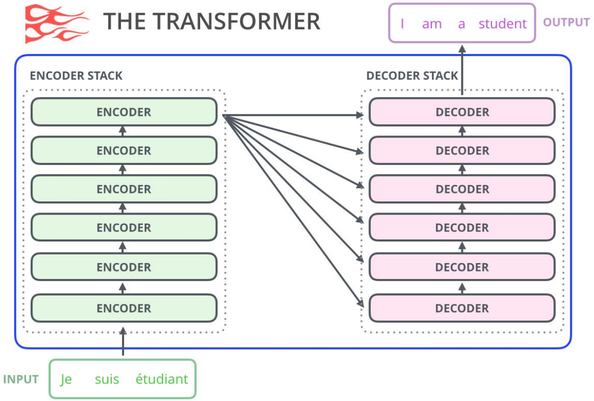
sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt

eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum. Lorem ipsum sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure

aliquip ex ea commodo consequat.

Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum. Lorem ipsum sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Lorem ipsum dolor sit amet, consectetur adipiscing

Model



Learning

Enter text:

One, two,

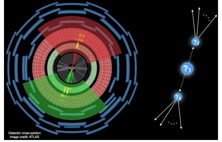
One , two ,

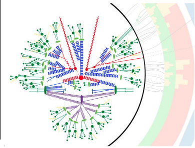
3198 11 734 11

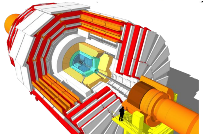
3198 11 734 11

Prediction

#	probs	next token ID	predicted next token
0	39.71%	1115	three
1	16.97%	290	and
2	7.55%	734	two



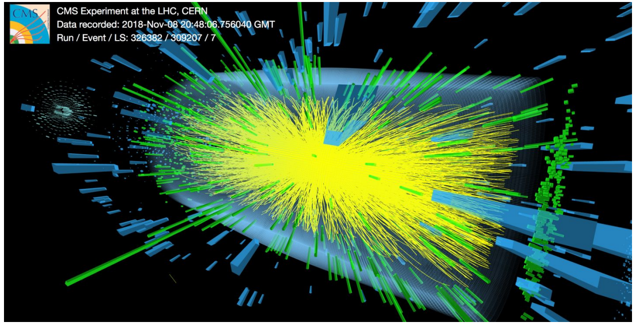


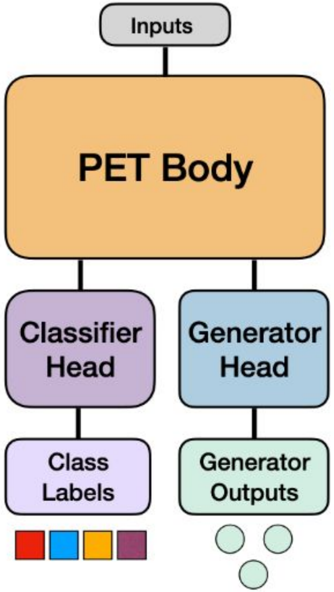


Data

Model

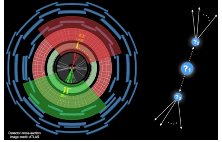
Learning

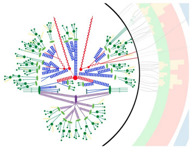


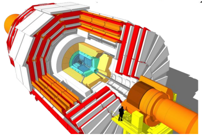


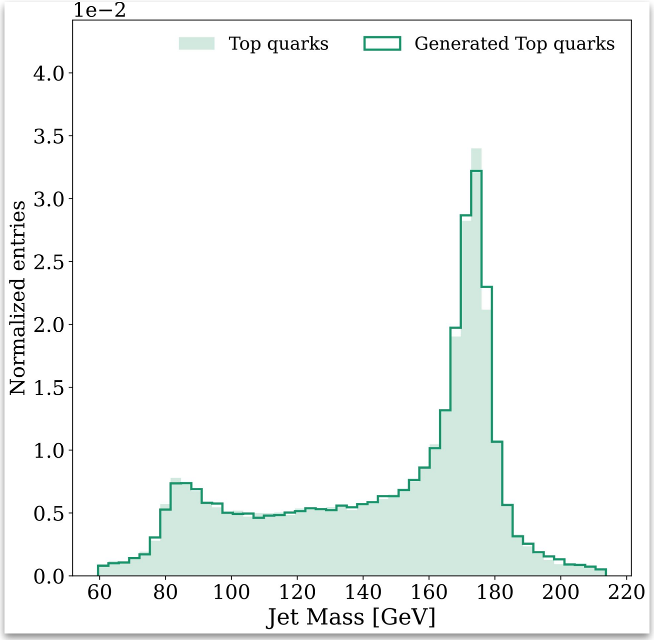
$$\begin{aligned}\mathcal{L} &= \mathcal{L}_{\text{class}} + \mathcal{L}_{\text{gen}} + \mathcal{L}_{\text{class smear}} \\ &= \text{CE}(y, y_{\text{pred}}) + \|\mathbf{v} - \mathbf{v}_{\text{pred}}\|^2 + \alpha^2 \text{CE}(y, \hat{y}_{\text{pred}})\end{aligned}$$

Improving Generative Models



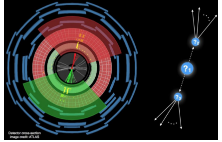


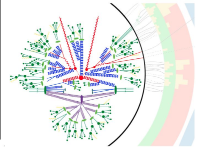


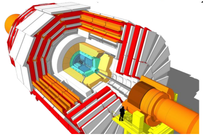


Jet class	Model	W_1^{PM} ($\times 10^{-3}$)	W_1^{P} ($\times 10^{-3}$)	W_1^{PEFP} ($\times 10^{-5}$)	FPND	Cov↑	MMD
Gluon	FPCD [52]	**0.36 ± 0.08**	**0.34 ± 0.09**	0.47 ± 0.13	0.07	0.55	0.03
	FPCD 1 [52]	0.65 ± 0.11	**0.34 ± 0.06**	0.60 ± 0.09	0.11	0.55	0.03
	MP-GAN [44]	0.69 ± 0.07	1.8 ± 0.2	0.9 ± 0.6	0.20	0.54	0.037
	EPiC-GAN [45]	**0.3 ± 0.1**	1.6 ± 0.2	0.4 ± 0.2	1.01 ± 0.07	-	-
	PET generator	0.42 ± 0.10	**0.36 ± 0.08**	**0.35 ± 0.08**	0.04	0.55	0.03
	PET generator (Ideal)	**0.36 ± 0.08**	**0.34 ± 0.09**	0.47 ± 0.13	0.07	0.55	0.03
	OmniLearn	**0.38 ± 0.08**	**0.33 ± 0.07**	**0.33 ± 0.09**	**0.02**	0.55	0.03
	OmniLearn (Ideal)	**0.33 ± 0.06**	**0.29 ± 0.08**	**0.30 ± 0.07**	**0.02**	0.55	0.03
Light Quark	FPCD [52]	0.52 ± 0.07	**0.27 ± 0.06**	0.38 ± 0.11	0.08	0.49	0.02
	FPCD 1 [52]	0.59 ± 0.08	0.36 ± 0.08	0.50 ± 0.08	0.09	0.48	0.02
	MP-GAN [44]	0.6 ± 0.2	4.9 ± 0.5	0.7 ± 0.4	0.35	0.50	0.026
	EPiC-GAN [45]	0.5 ± 0.1	4.0 ± 0.4	0.8 ± 0.4	0.43 ± 0.03	-	-
	PET generator	0.39 ± 0.12	0.35 ± 0.06	**0.24 ± 0.10**	0.03	**0.54**	0.02
	PET generator (Ideal)	0.31 ± 0.08	0.38 ± 0.10	**0.23 ± 0.07**	0.03	0.53	0.02
	OmniLearn	**0.24 ± 0.03**	**0.32 ± 0.07**	**0.24 ± 0.08**	0.02	**0.54**	0.02
	OmniLearn (Ideal)	0.31 ± 0.08	**0.30 ± 0.09**	**0.26 ± 0.08**	**0.01**	**0.54**	0.02
Top Quark	FPCD [52]	0.51 ± 0.07	0.41 ± 0.12	1.25 ± 0.19	0.17	0.58	0.05
	FPCD 1 [52]	1.22 ± 0.09	0.46 ± 0.10	2.66 ± 0.26	0.56	0.57	0.05
	MP-GAN [44]	0.6 ± 0.2	2.3 ± 0.3	2 ± 1	0.37	0.57	0.071
	EPiC-GAN [45]	0.5 ± 0.1	2.1 ± 0.1	1.7 ± 0.3	0.31 ± 0.037	-	-
	PET generator	0.44 ± 0.03	**0.29 ± 0.07**	**1.09 ± 0.23**	0.07	0.58	0.05
	PET generator (Ideal)	**0.41 ± 0.07**	**0.34 ± 0.08**	**1.22 ± 0.23**	0.07	0.58	0.05
	OmniLearn	0.43 ± 0.06	**0.30 ± 0.07**	1.31 ± 0.18	0.04	0.58	0.05
	OmniLearn (Ideal)	**0.36 ± 0.05**	0.41 ± 0.08	**1.02 ± 0.20**	**0.03**	0.58	0.05
W Boson	FPCD [52]	0.26 ± 0.03	0.39 ± 0.08	0.15 ± 0.02	-	0.56	0.02
	FPCD 1 [52]	0.94 ± 0.06	0.42 ± 0.09	0.35 ± 0.03	-	0.56	0.02
	PET generator	**0.17 ± 0.04**	**0.26 ± 0.05**	**0.11 ± 0.02**	-	0.56	0.02
	PET generator (Ideal)	**0.15 ± 0.02**	**0.31 ± 0.07**	**0.12 ± 0.03**	-	**0.57**	0.02
	OmniLearn	**0.19 ± 0.03**	**0.27 ± 0.07**	**0.10 ± 0.02**	-	**0.57**	0.02
	OmniLearn (Ideal)	**0.16 ± 0.06**	**0.28 ± 0.04**	**0.10 ± 0.02**	-	**0.57**	0.02
Z Boson	FPCD [52]	**0.21 ± 0.04**	0.40 ± 0.13	0.18 ± 0.03	-	0.56	0.02
	FPCD 1 [52]	0.99 ± 0.05	0.35 ± 0.06	0.49 ± 0.03	-	0.56	0.02
	PET generator	**0.22 ± 0.04**	**0.32 ± 0.07**	0.20 ± 0.04	-	**0.57**	0.02
	PET generator (Ideal)	**0.18 ± 0.10**	**0.30 ± 0.08**	**0.14 ± 0.02**	-	0.56	0.02
	OmniLearn	**0.19 ± 0.07**	**0.32 ± 0.09**	**0.12 ± 0.03**	-	**0.57**	0.02
	OmniLearn (Ideal)	**0.22 ± 0.05**	**0.27 ± 0.06**	**0.13 ± 0.02**	-	**0.57**	0.02

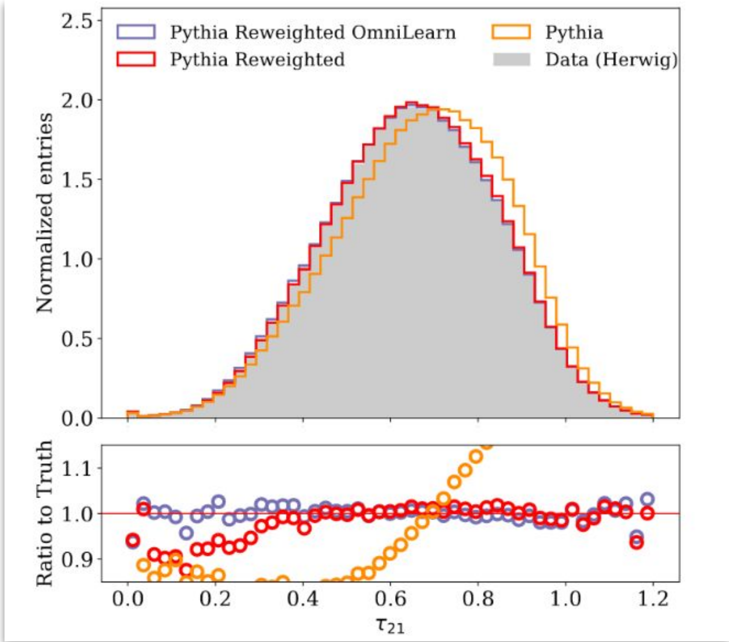
Simultaneously improving **generative models**, unfolding, and anomaly detection in 9 different benchmarks!







Improving Unfolding

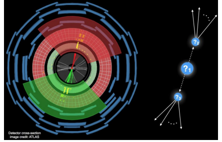


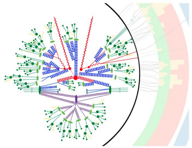
Metric	MultiFold	UniFold	IBU	OmniFold		
				DeepSets	PET classifier	OmniLearn
Jet mass	3.80	8.82	9.31	**2.77**	**2.8±0.9**	**2.6±0.8**
N	0.89	1.46	1.51	**0.33**	0.50±0.15	**0.34±0.1**
Jet Width	0.09	0.15	0.11	0.10	0.09±0.02	**0.07±0.01**
$\log \rho$	0.37	0.59	0.71	0.35	0.23±0.07	**0.14±0.03**
τ_{21}	0.26	1.11	1.10	0.53	0.13±0.03	**0.05±0.01**
z_g	**0.15**	0.59	0.37	0.68	0.19±0.03	0.21±0.04

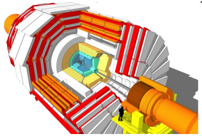
Training time reduced by a **factor 2**!

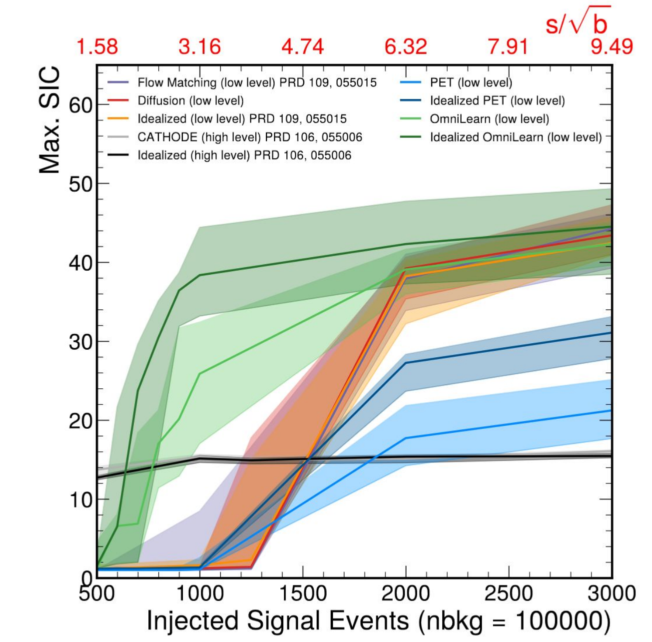
Simultaneously improving generative models, **unfolding**, and anomaly detection in 9 different benchmarks!

Improving Anomaly Detection





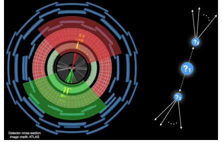


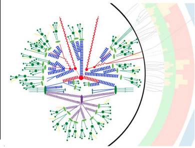


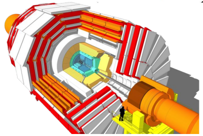
Improved sensitivity to new physics: requires **4 times** less data to find the signal!

Simultaneously improving generative models, unfolding, and **anomaly detection** in 9 different benchmarks!

Improving Everything!







	Acc	AUC	$1/\epsilon_B$	
			$\epsilon_S = 0.5$	$\epsilon_S = 0.3$
P-CNN [37]	0.827	0.9002	34.7	91.0
PFN [34]	-	0.9005	34.7±0.4	-
ParticleNet [37]	0.840	0.9116	39.8±0.2	98.6±1.3
rPCN [38]	-	0.9081	38.6 ± 0.5	-
ParT [41]	0.840	0.9121	41.3 ± 0.3	101.2 ± 1.1
ParT-f.t. [41]	0.843	0.9151	42.4 ± 0.2	**107.9 ± 0.5**
PET classifier	0.837	0.9110	39.92±0.1	104.9 ± 1.5
OmniLearn	**0.844**	**0.9159**	**43.7±0.3**	**107.7 ± 1.5**

	AUC	Acc	$1/\epsilon_B$	
			$\epsilon_S = 0.5$	$\epsilon_S = 0.8$
ResNet 50	0.885	0.803	21.4	5.13
EFN	0.901	0.819	26.6	6.12
hlDNN	0.938	0.863	51.5	10.5
DNN	0.942	0.868	67.7	12.0
PFN	0.954	0.882	108.0	15.9
ParticleNet	0.961	0.894	153.7	20.4
PET classifier (4M)	0.959	0.890	146.5	19.4
OmniLearn (4M)	0.961	0.894	172.1	20.8
PET classifier (40M)	0.964	0.898	201.4	23.6
OmniLearn (40M)	**0.965**	**0.899**	**207.30**	**24.10**

More benchmarks in
V. Mikuni, B. Nachman,
arXiv:2404.16091

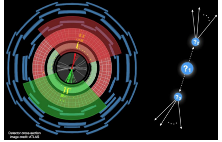
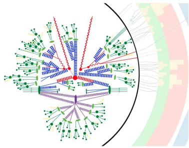
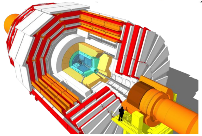
	AUC	Acc	$1/\epsilon_B$	
			$\epsilon_S = 0.5$	$\epsilon_S = 0.8$
PET classifier	0.875	0.796	23.91 ± 0.07	4.770 ± 0.001
OmniLearn	**0.877**	**0.797**	**24.36 ± 0.01**	**4.836 ± 0.004**

Jet class	Model	W_1^{PM} ($\times 10^{-3}$)	W_1^{P} ($\times 10^{-3}$)	W_1^{PEFP} ($\times 10^{-5}$)	FPND	Cov↑	MMD
Gluon	FPCD [52]	**0.36 ± 0.08**	**0.34 ± 0.09**	0.47 ± 0.13	0.07	0.55	0.03
	FPCD 1 [52]	0.65 ± 0.11	**0.34 ± 0.06**	0.60 ± 0.09	0.11	0.55	0.03
	MP-GAN [44]	0.69 ± 0.07	1.8 ± 0.2	0.9 ± 0.6	0.20	0.54	0.037
	EPiC-GAN [45]	**0.3 ± 0.1**	1.6 ± 0.2	0.4 ± 0.2	1.01 ± 0.07	-	-
	PET generator	0.42 ± 0.10	**0.36 ± 0.08**	**0.35 ± 0.08**	0.04	0.55	0.03
	PET generator (Ideal)	**0.36 ± 0.08**	**0.34 ± 0.09**	0.47 ± 0.13	0.07	0.55	0.03
	OmniLearn	**0.38 ± 0.08**	**0.33 ± 0.07**	**0.33 ± 0.09**	**0.02**	0.55	0.03
	OmniLearn (Ideal)	**0.33 ± 0.06**	**0.29 ± 0.08**	**0.30 ± 0.07**	**0.02**	0.55	0.03

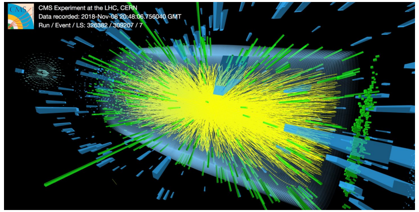
	PET classifier	OmniLearn
Jet mass	0.13±0.03	**0.027±0.008**
N	0.13±0.03	**0.05±0.02**
Jet Width	0.09±0.02	**0.02±0.01**
$\log \rho$	0.08±0.02	**0.03±0.01**
τ_{21}	0.08±0.03	**0.02±0.01**
z_g	0.04±0.01	**0.001±0.004**

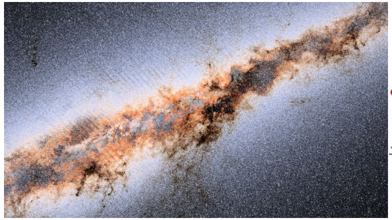
Simultaneously improving **generative models**, **unfolding**, and **anomaly detection** in **9 different benchmarks**!

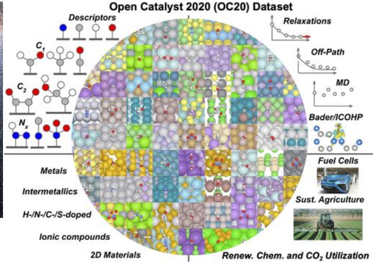
Future

Data







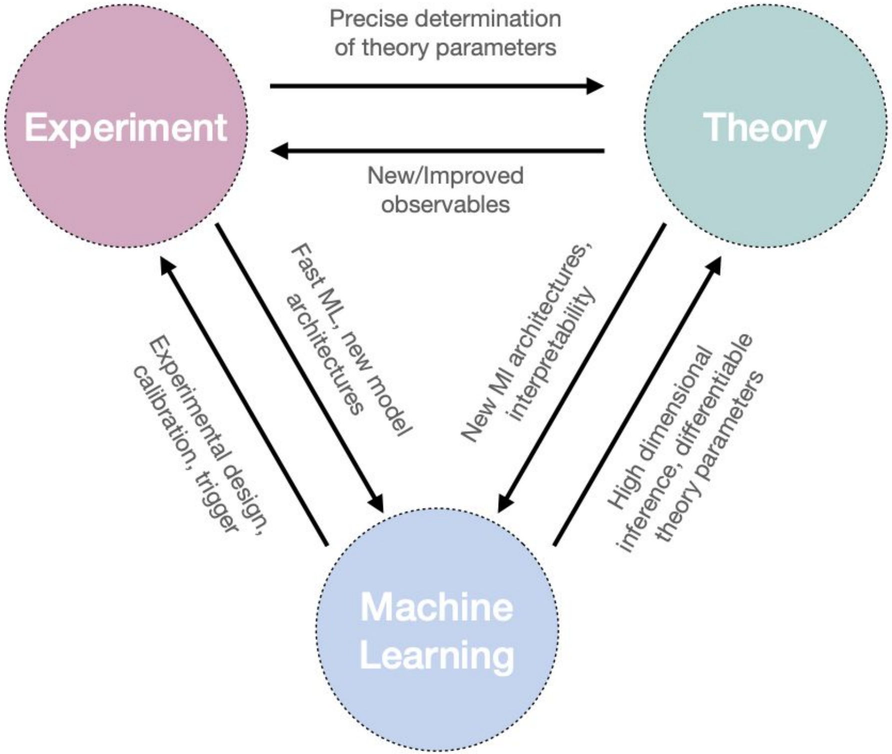
Model

?

Learning

?

Conclusions



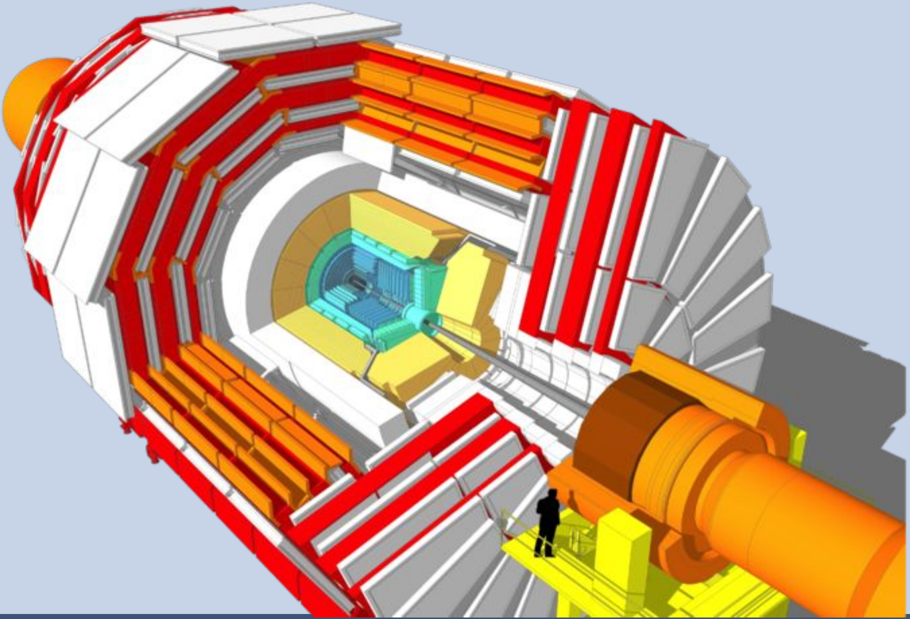
AI is revolutionizing the way to do science

- At the LHC, large amounts of data motivate the use of AI to accelerate discovery
- Beyond the LHC, AI enables interdisciplinary research
- **Interdisciplinary models** could bring new discoveries: Foundational models for Science!

THANKS!

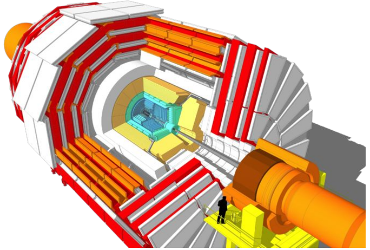
Any questions?

Backup



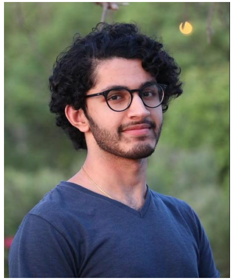
Fast Detector Simulation

Improving Simulations even Further



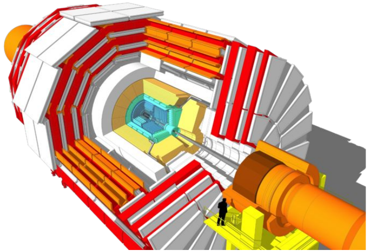
Diffusion generative models in latent space for calorimeter simulation.
Thandikire Madula, PhD Student at UCL: submission accepted at **NeurIPS 2023 ML4PS Workshop**

Calorimeter detector simulation with continuous normalizing flows.
Chirag Furia, CS Undergraduate Student at Brown University: submission accepted at **NeurIPS 2023 ML4PS Workshop**



Faster diffusion generative models for jet generation.
Yash Melkani, Physics Undergraduate Student at UC Berkeley

Diffusion Generative Models



ON THE THEORY OF THE BROWNIAN MOTION

By G. E. Uhlenbeck and L. S. Ornstein

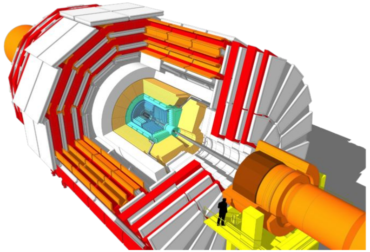
University of Michigan, Ann Arbor and Physisch Laboratorium der R. U. Utrecht, Holland

(Received July 7, 1930)

Abstract

With a method first indicated by Ornstein the mean values of *all* the powers of the velocity u and the displacement s of a free particle in Brownian motion are calculated. It is shown that $u-u_0\exp(-\beta t)$ and $s-u_0/\beta[1-\exp(-\beta t)]$ where u_0 is the initial velocity and β the friction coefficient divided by the mass of the particle, follow the normal Gaussian distribution law. For s this gives the exact frequency distribution corresponding to the exact formula for $\overline{s^2}$ of Ornstein and Fürth. Discussion is given of the connection with the Fokker-Planck partial differential equation. By the same method exact expressions are obtained for the square of the deviation of a harmonically bound particle in Brownian motion as a function of the time and the initial deviation. Here the periodic, aperiodic and overdamped cases have to be treated separately. In the last case, when β is much larger than the frequency and for values of $t \gg \beta^{-1}$, the formula takes the form of that previously given by Smoluchowski.

Diffusion Generative Models



A Mechanical Model of Brownian Motion

Communications in Mathematical Physics

D. Dürr*, S. Goldstein**, and J. L. Lebowitz***

Department of Mathematics, Rutgers University, New Brunswick, NJ 08903, USA

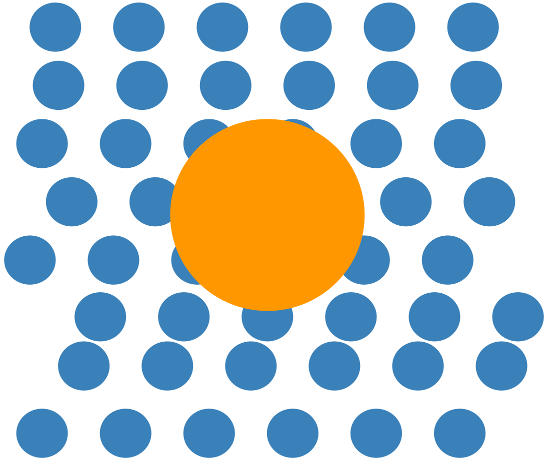
Abstract. We consider a dynamical system consisting of one large massive particle and an infinite number of light point particles. We prove that the motion of the massive particle is, in a suitable limit, described by the Ornstein-Uhlenbeck process. This extends to three dimensions previous results by Holley in one dimension.

The ultimate mathematical idealization of this phenomenon is the Ornstein-Uhlenbeck process for the position and velocity of the Brownian particle $(\underline{X}_t, \underline{V}_t)$, described by the stochastic differential equations

$$d\underline{X}_t = \underline{V}_t dt\,, \tag{0.1}$$

$$d\underline{V}_t = -a\underline{V}_t dt + \sqrt{D}\, d\underline{W}_t\,, \quad a \geqq 0\,, \quad D \geqq 0\,, \quad \underline{W}_t = \text{Wiener process}. \tag{0.2}$$

The position process $\underline{X}_t$ converges in an appropriate limit (e.g. $a \to \infty$, $a^2/D = \text{const}$) to a Wiener process.



Score matching/denoising/diffusion

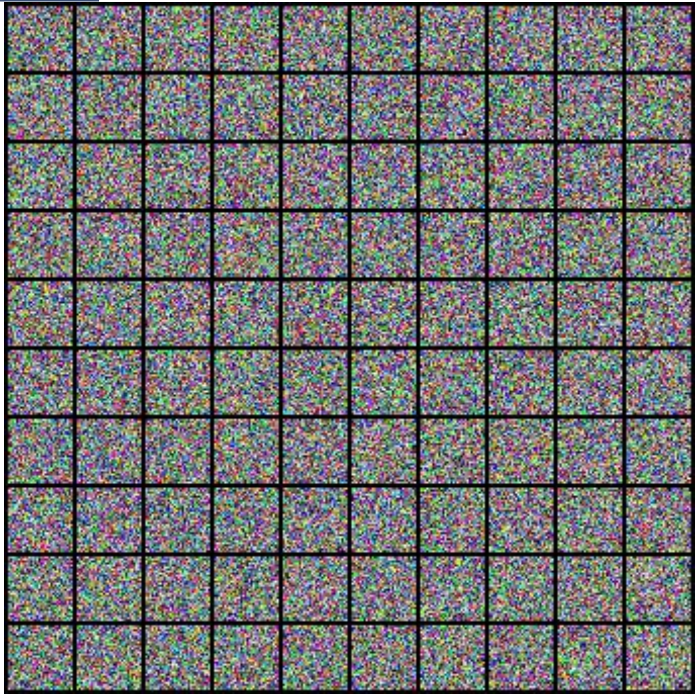
Denoise diffusion models are the newest state-of-the-art generative models for image generation.

Pros:

- **Stable training**: convex loss function
- **Scalability**: Network complexity is more sensitive to the architecture than the dimensionality
- **Access to data likelihood after training**: similar to NFs, but overall normalization is not required during training

Cons:

- **Slow sampling**: Possibly **1000s** of model evaluations to generate realistic images



Score-matching

- The common choice for **λ(t) is σ(t)²** resulting in the loss function

$$\frac{1}{2}\mathbb{E}_t\mathbb{E}_{p_t(\tilde{x})}\left[\|\sigma(t)s_\theta(\tilde{x},t)+\epsilon(0,1)\|_2^2\right]$$

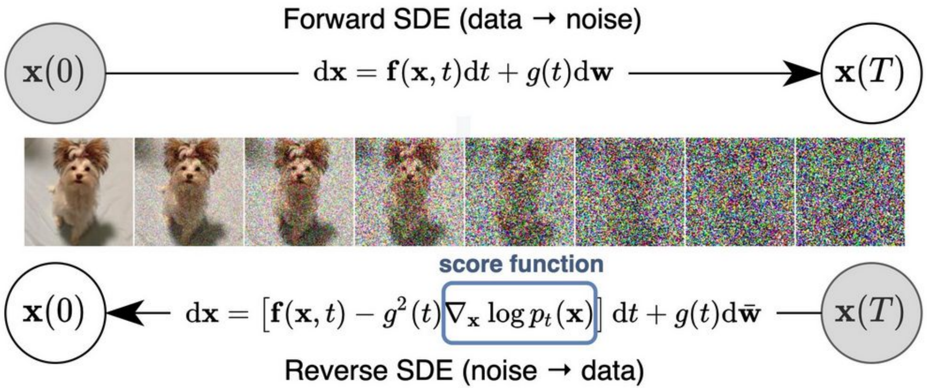
- Another important result is when **λ(t) is g(t)²** that represents an [upper bound of the data likelihood]()

$$\mathrm{KL}(p_0(\mathbf{x})\|p_\theta(\mathbf{x})) \leq \frac{T}{2}\mathbb{E}_{t\in\mathcal{U}(0,T)}\mathbb{E}_{p_t(\mathbf{x})}[\lambda(t)\|\nabla_\mathbf{x}\log p_t(\mathbf{x})-\mathbf{s}_\theta(\mathbf{x},t)\|_2^2] + \mathrm{KL}(p_T \parallel \pi).$$

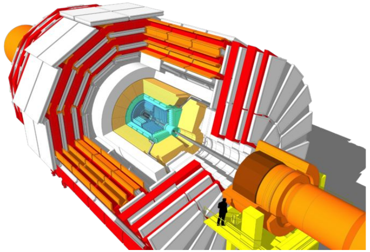
- Allowing the **maximum-likelihood** training of diffusion models!

Generation

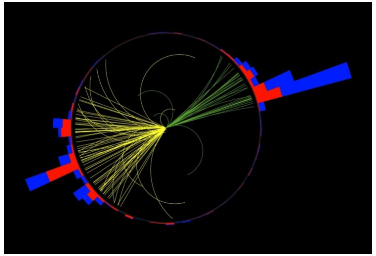
- Generation of new samples is done by solving the **reverse SDE**
- Langevin dynamics is used to draw samples from **p(x)** using only the **score function**
- High fidelity samples require small time steps,
- For Calorimeter generation, **O(100)** evaluations are enough to produce precise results

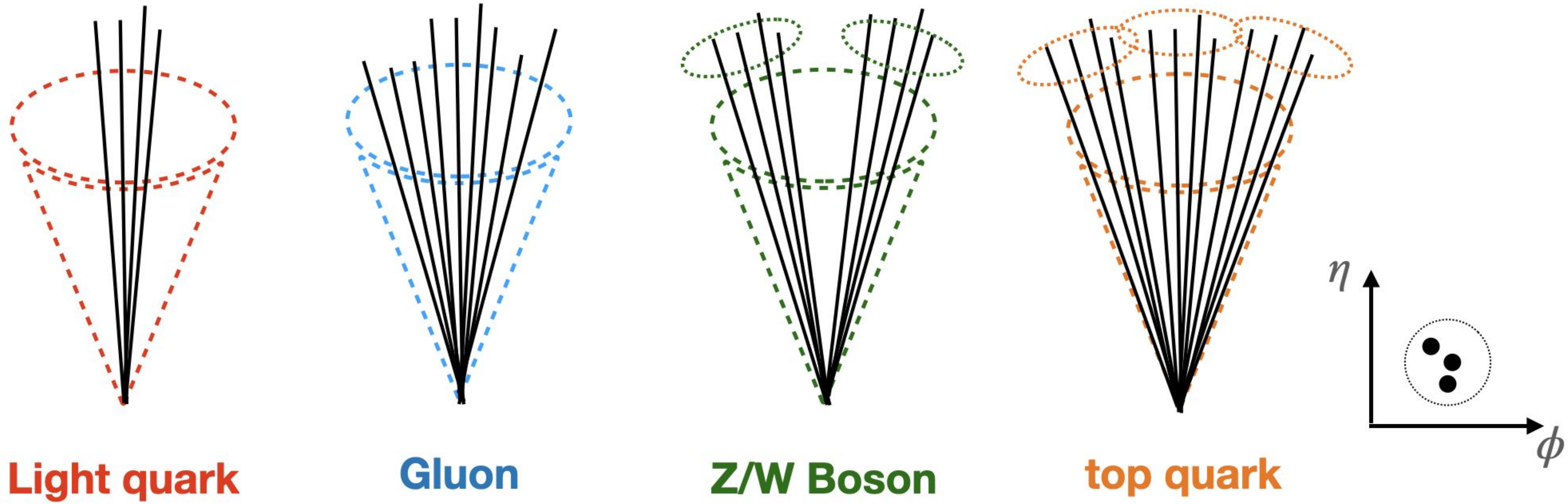


$$\mathbf{x}_{i+1} \leftarrow \mathbf{x}_i + \epsilon \nabla_{\mathbf{x}} \log p(\mathbf{x}) + \sqrt{2\epsilon}\, \mathbf{z}_i, \quad i = 0, 1, \cdots, K,$$



Particle generation

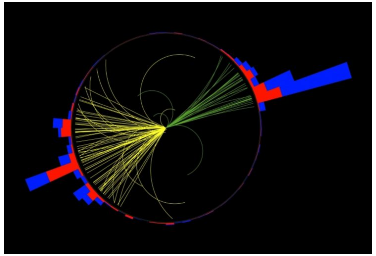




JetNet30 and JetNet150 Datasets

- Datasets with up to **30** or **150** particles
- Multiple jet classes including: **Top quarks, W/Z bosons, Light quarks, Gluons**

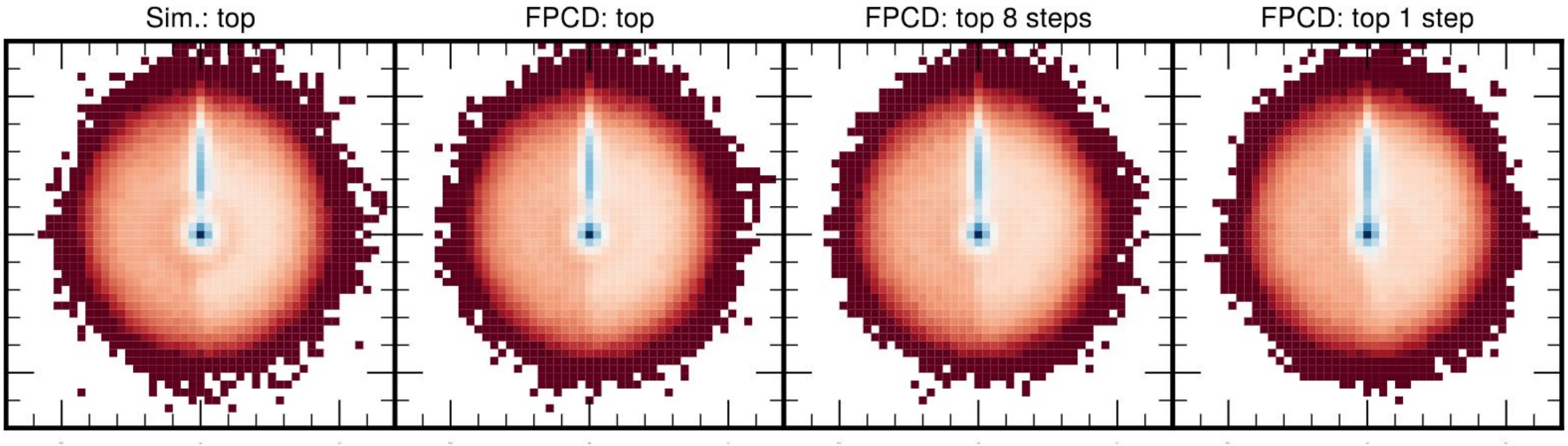
Particle generation



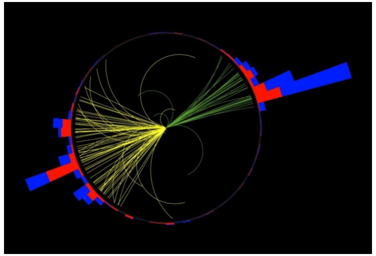
Mikuni, V, Nachman, B., and M. Pettee *Phys. Rev. D 108, 036025*

Progressive distillation is used to reduce the overall number of function evaluations

- **Single-shot** generation with almost no performance degradation



Results

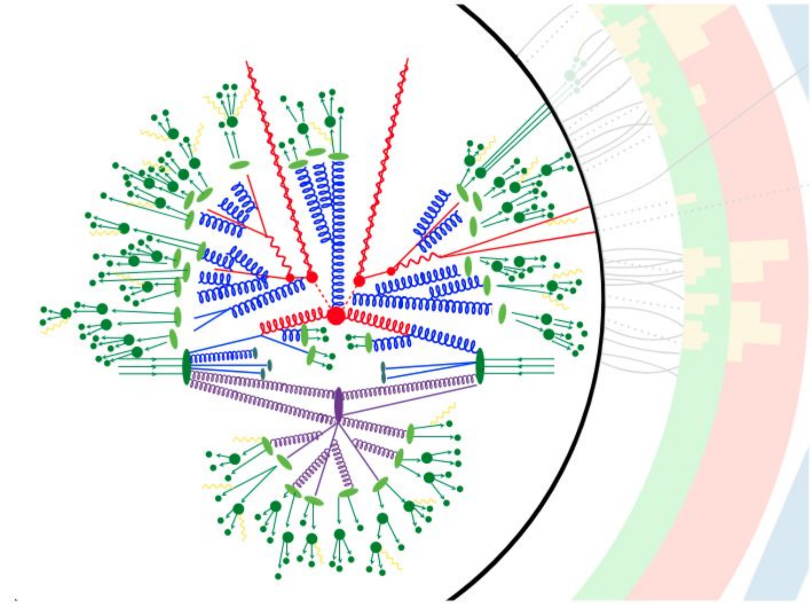


Mikuni, V, Nachman, B., and M. Pettee *Phys. Rev. D 108, 036025*

Jet class	Model	W_1^{PM} ($\times10^{-3}$)	W_1^{P} ($\times10^{-3}$)	W_1^{PEFP} ($\times10^{-5}$)	FPND	Cov↑	MMD
Gluon	FPCD	**0.36 ± 0.08**	**0.34 ± 0.09**	**0.47 ± 0.13**	**0.07**	**0.55**	**0.03**
	FPCD 8	0.60 ± 0.16	**0.36 ± 0.07**	**0.54 ± 0.09**	**0.07**	**0.55**	**0.03**
	FPCD 1	0.65 ± 0.11	**0.34 ± 0.06**	0.60 ± 0.09	0.11	**0.55**	**0.03**
	MP-GAN [35]	0.69 ± 0.07	1.8 ± 0.2	0.9 ± 0.6	0.20	0.54	0.037
	EPiC-GAN [36]	**0.3 ± 0.1**	1.6 ± 0.2	**0.4 ± 0.2**	1.01 ± 0.07	-	-
Light Quark	FPCD	**0.52 ± 0.07**	**0.27 ± 0.06**	**0.38 ± 0.11**	**0.08**	0.49	**0.02**
	FPCD 8	**0.59 ± 0.14**	0.35 ± 0.05	**0.44 ± 0.07**	0.09	0.48	**0.02**
	FPCD 1	**0.59 ± 0.08**	0.36 ± 0.08	0.50 ± 0.08	0.09	0.48	**0.02**
	MP-GAN [35]	**0.6 ± 0.2**	4.9 ± 0.5	0.7 ± 0.4	0.35	**0.50**	0.026
	EPiC-GAN [36]	**0.5 ± 0.1**	4.0 ± 0.4	0.8 ± 0.4	0.43 ± 0.03	-	-
Top Quark	FPCD	**0.51 ± 0.07**	**0.41 ± 0.12**	**1.25 ± 0.19**	**0.17**	**0.58**	**0.05**
	FPCD 8	0.80 ± 0.06	**0.45 ± 0.12**	1.91 ± 0.30	0.37	**0.58**	**0.05**
	FPCD 1	1.22 ± 0.09	**0.46 ± 0.10**	2.66 ± 0.26	0.56	0.57	**0.05**
	MP-GAN [35]	**0.6 ± 0.2**	2.3 ± 0.3	2 ± 1	0.37	0.57	0.071
	EPiC-GAN [36]	**0.5 ± 0.1**	2.1 ± 0.1	1.7 ± 0.3	0.31 ± 0.037	-	-
W Boson	FPCD	**0.26 ± 0.03**	**0.39 ± 0.08**	**0.15 ± 0.02**	-	**0.56**	**0.02**
	FPCD 8	0.48 ± 0.04	**0.38 ± 0.05**	0.22 ± 0.02	-	0.55	**0.02**
	FPCD 1	0.94 ± 0.06	**0.42 ± 0.09**	0.35 ± 0.03	-	**0.56**	**0.02**
Z Boson	FPCD	**0.21 ± 0.04**	**0.40 ± 0.13**	**0.18 ± 0.03**	-	**0.56**	**0.02**
	FPCD 8	0.40 ± 0.04	**0.35 ± 0.04**	0.27 ± 0.03	-	**0.56**	**0.02**
	FPCD 1	0.99 ± 0.05	**0.35 ± 0.06**	0.49 ± 0.03	-	**0.56**	**0.02**

Multiple physics inspire metrics used to evaluate the performance of the generative model, achieving **SOTA** in many categories

- Single-shot model is still performant and **1000** times faster than the full simulation

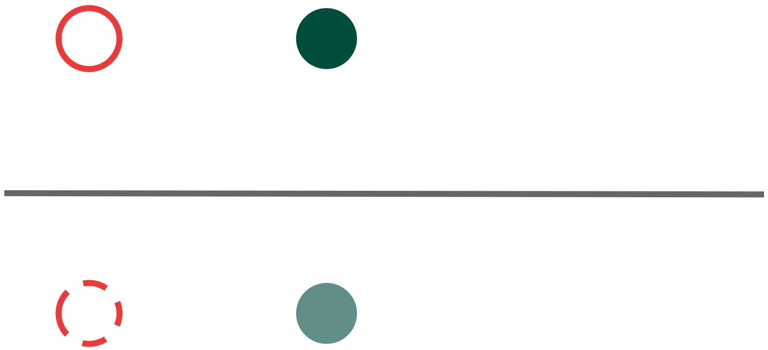


Multidimensional Unfolding

Omnifold

Reco level

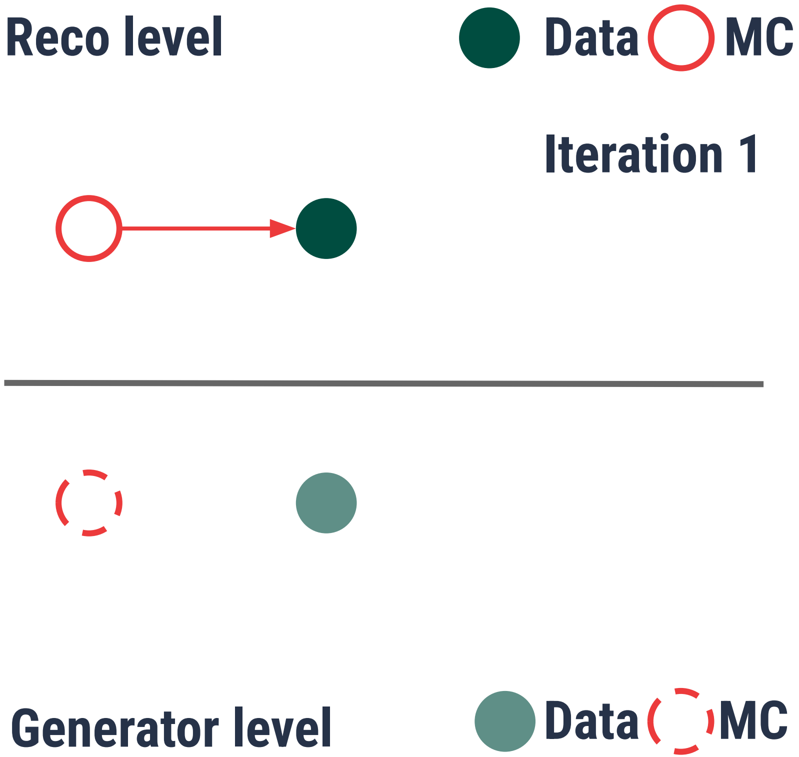
Data MC



Generator level

Data MC

Omnifold



Step 1:

- Train a classifier to separate **data** from **MC** events
- Reweight **reco level MC** with weights:

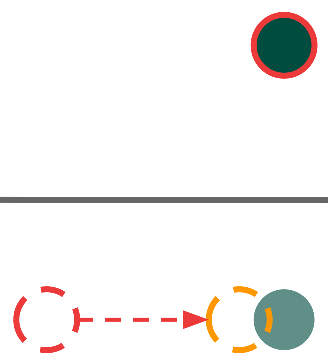
$$W(\text{reco}) = p_{\text{Data}}(\text{reco})/p_{\text{MC}}(\text{reco})$$

Omnifold

Reco level

Data MC

Iteration 1



Generator level

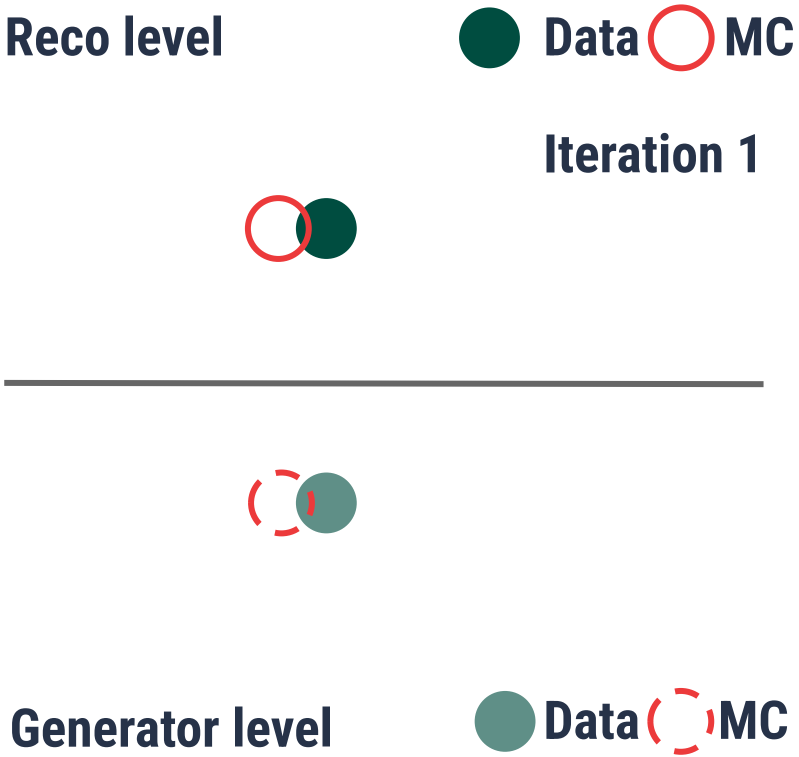
Data MC MC reweighted

Step 2:

- **Pull weights** from **step 1** to generator level events
- Train a classifier to separate **initial MC at gen level** from **reweighted MC** events
- Define a **new simulation** with weights that are a **proper function of gen level kinematics**

$$W(gen) = p_{weighted\ MC}(gen)/p_{MC}(gen)$$

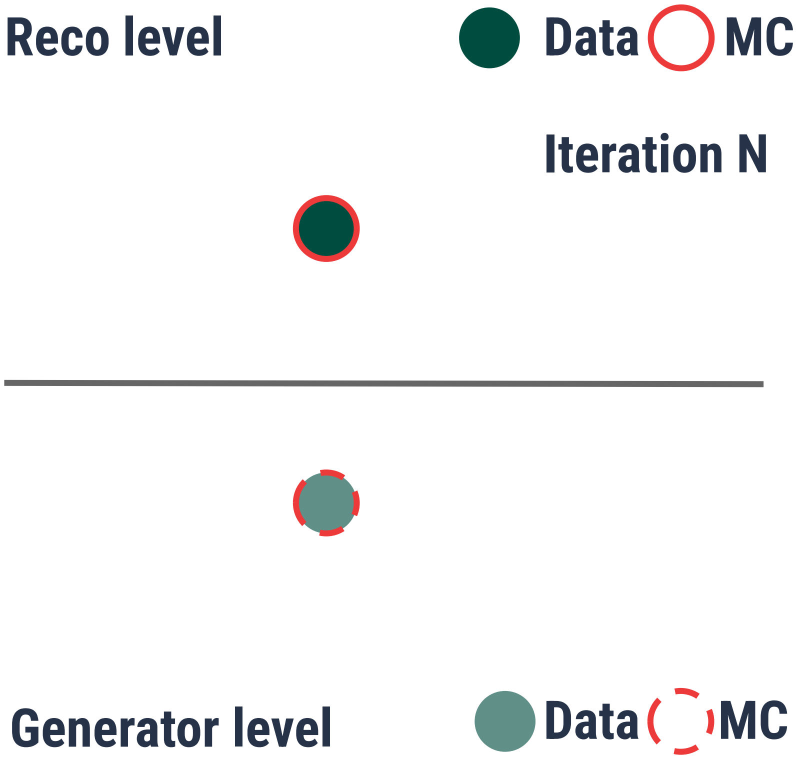
Omnifold



Start again from **step 1** using the **new simulation** after **pushing** the weights from **step 2**

- Guaranteed convergence to the maximum likelihood estimate of the generator-level distribution when number of iterations go to infinite
- In practice, less than 10 iterations are enough to achieve convergence

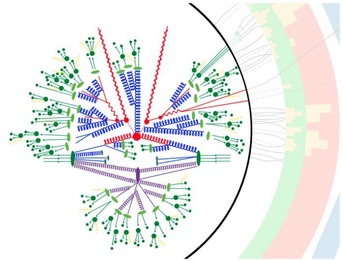
Omnifold

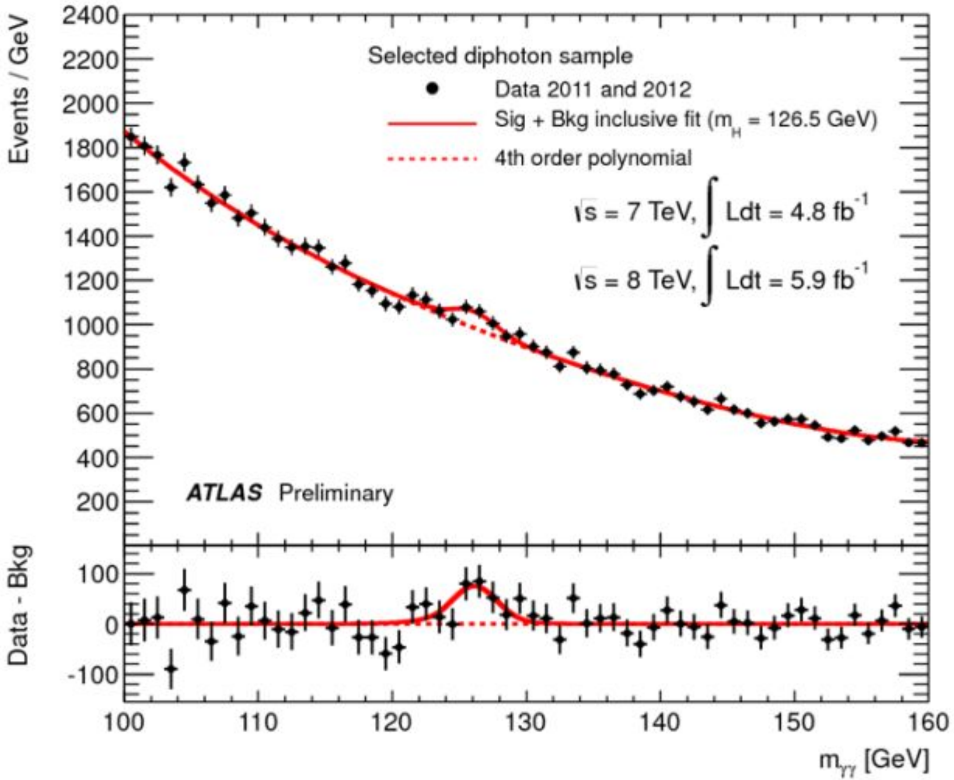


Start again from **step 1** using the **new simulation** after **pushing** the weights from **step 2**

- **Guaranteed convergence** to the maximum likelihood estimate of the generator-level distribution when number of iterations goes to infinite
- In practice, **less than 10 iterations** are enough to achieve convergence

The Challenge

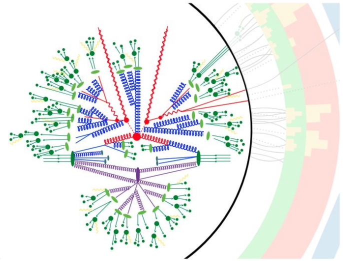


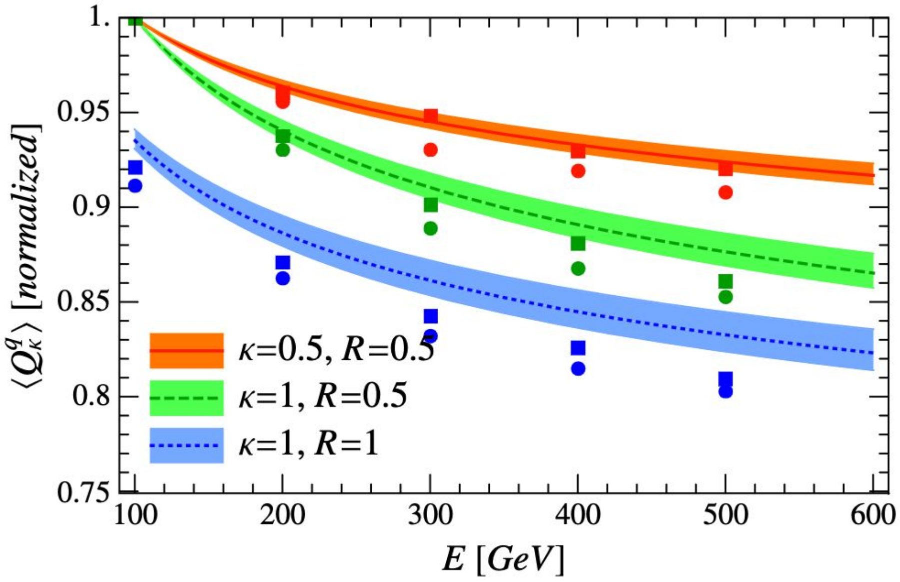


Almost all measurements at the LHC are reported as **histograms**

- Natural representation for counting problems
- Well-understood properties

The Challenge





Not everything is naturally represented by a histogram!

- **Moments of distributions**: theory is only sensitive to moments of observables

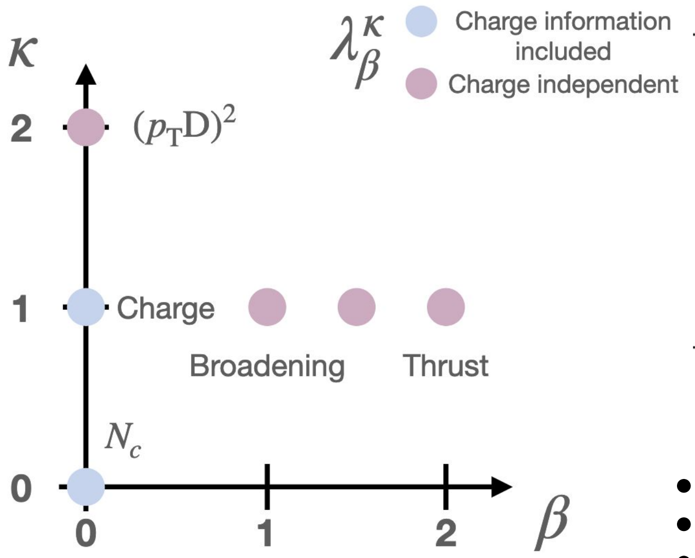
$$\langle \mathcal{Q}_\kappa^q \rangle = \frac{1}{16\pi^3} \frac{\widetilde{\mathcal{J}}_{qq}(E, R, \kappa, \mu)}{\mathcal{J}_q(E, R, \mu)} \sum_h Q_h \widetilde{D}_q^h(\kappa, \mu)$$

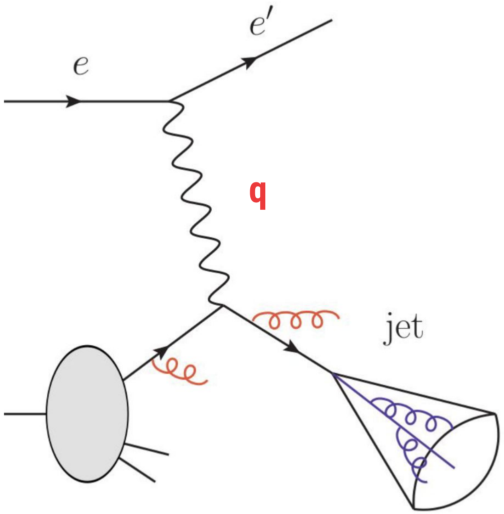
D. Krohn, M. D. Schwartz, T. Lin, and W. J. Waalewijn Phys. Rev. Lett. **110**, 212001

Jet angularities

Use jet observables to study different properties of QCD physics:

- Infrared and collinear (IRC) safe λ^1_a, a = [0,0.5,1] and unsafe $\mathbf{p_T D}$ angularities
- Charge dependent observables: $\mathbf{Q_j}$ and $\mathbf{N_c}$
- Study the evolution of the observables with energy scale $\mathbf{Q^2 = -q^2}$



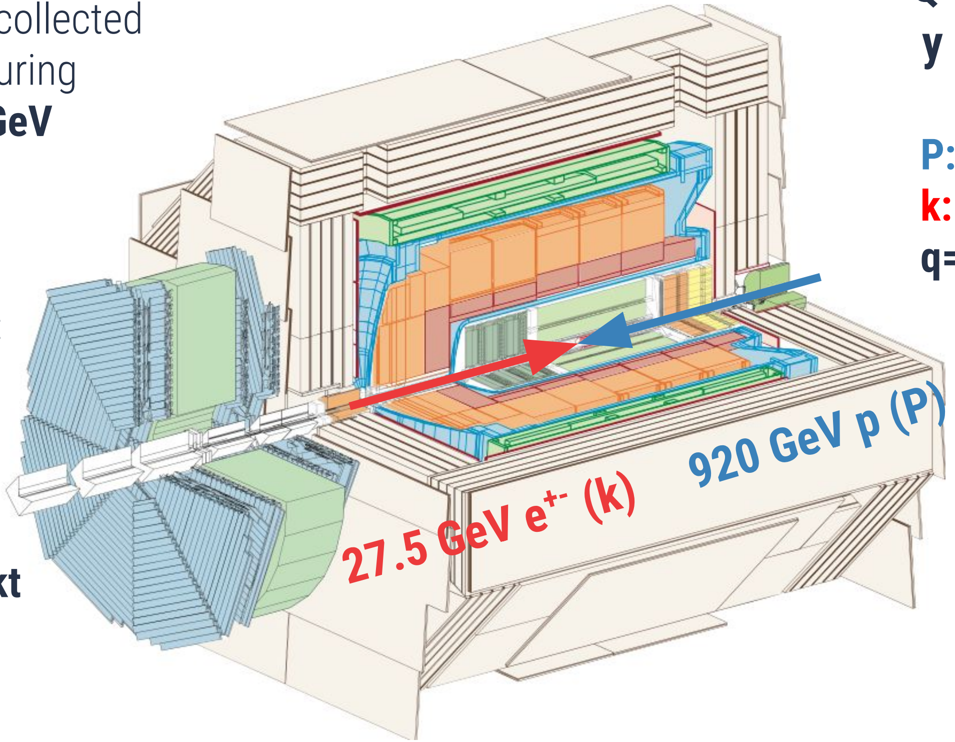


- z_i: longitudinal momentum fraction
- q_i: charge
- R_i distance from jet axis in (eta,phi)

$$\lambda^\kappa_\beta = \sum_{i\in\text{jet}} z_i^\kappa \left(\frac{R_i}{R_0}\right)^\beta \qquad \tilde{\lambda}^\kappa_0 = Q_\kappa = \sum_{i\in\text{jet}} q_i \times z_i^\kappa .$$

Experimental setup

Using **228 pb^{-1}** of data collected by the **H1 Experiment** during **2006** and **2007** at **318 GeV center-of-mass energy**



$Q^2 = - q^2$

$y = Pq / pk$

P: incoming proton 4-vector
k: incoming electron 4-vector
q=k-k' : 4-momentum transfer

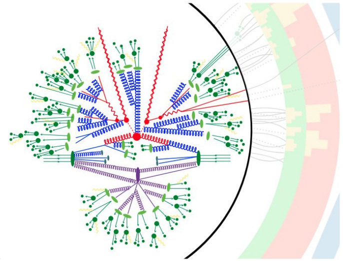
Phase space definition:

- $0.2 < y < 0.7$
- $Q^2 > 150\ GeV^2$
- Jet $p_T > 10$ GeV
- $-1 < \eta_{lab} < 2.5$

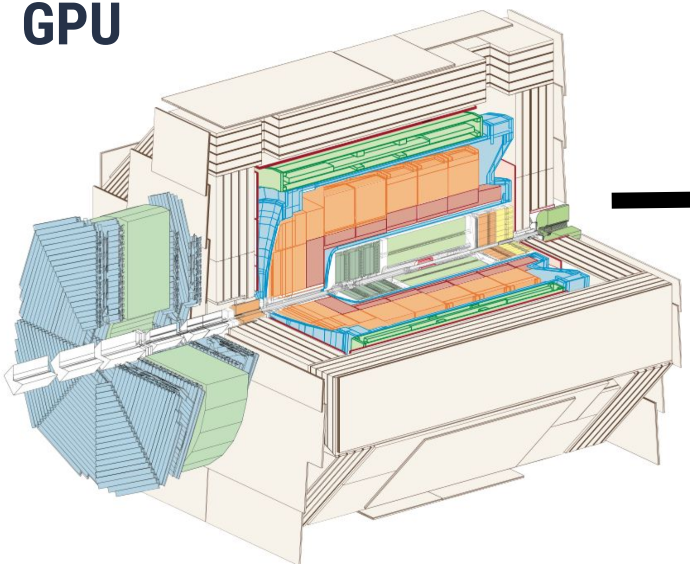
Jets are clustered with **kt** algorithm with **R=1.0**

Reconstructed hadrons using combined detector information: **energy flow algorithm**

Large Scale Computing

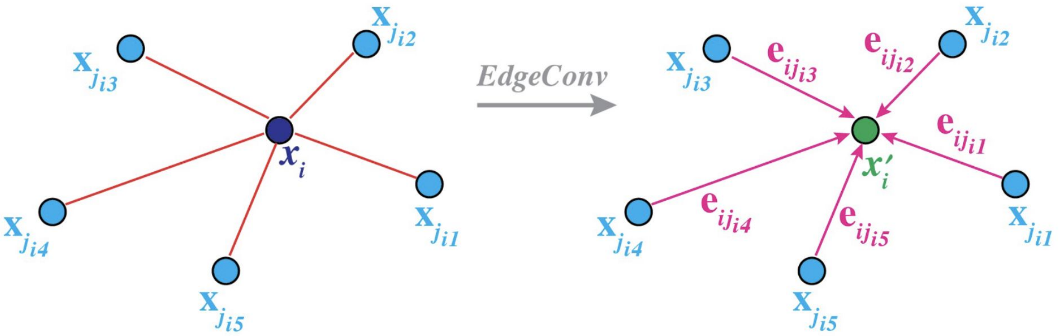


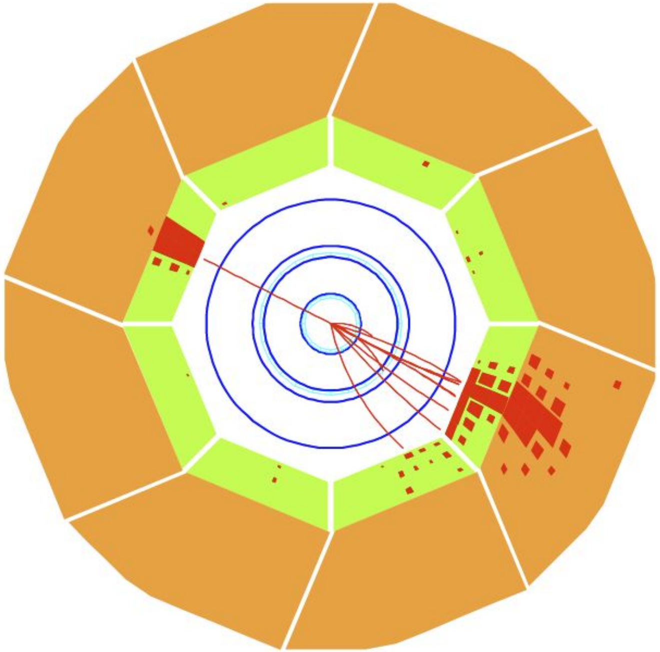
- **2800 neural networks** were trained to determine the final measurement
- One of the first uses of the **Perlmutter supercomputer** for science!
- Training with **128 GPUs simultaneously** while evaluation requires a **single GPU**



Extracting particle information

- Particle information is extracted using a **Point cloud transformer*** model
- Model takes **kinematic properties** of particles and use the distance between particles in η-φ to learn the relationship between particles
- Built in symmetries: **permutation invariance**
- Consider up to **30** particles per jet

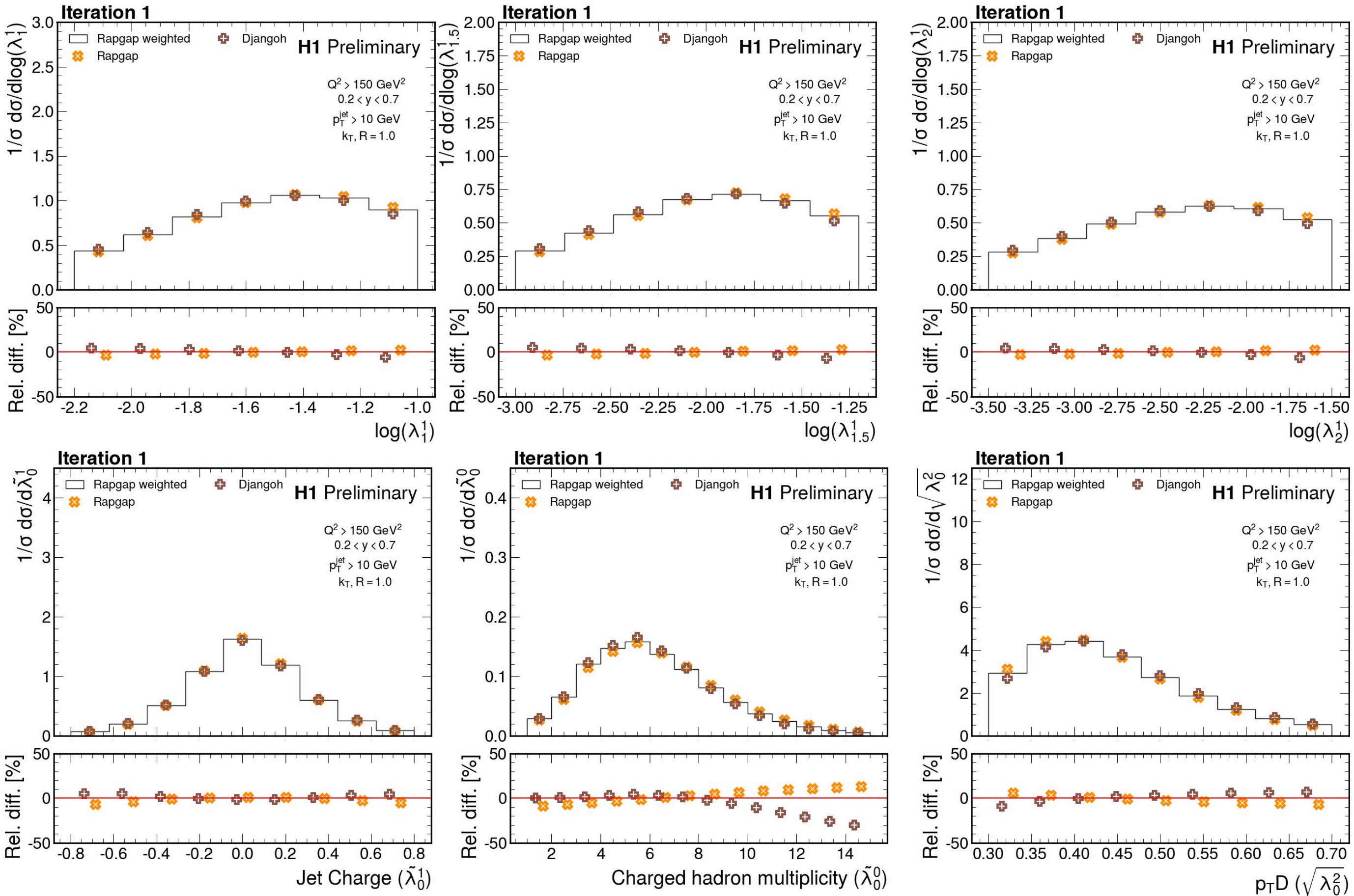




* V. Mikuni and F. Canelli 2021 *Mach. Learn.: Sci. Technol.* **2** 035027

Closure test

All distributions are unfolded **simultaneously without binning and without jet substructure information used at reco level!**

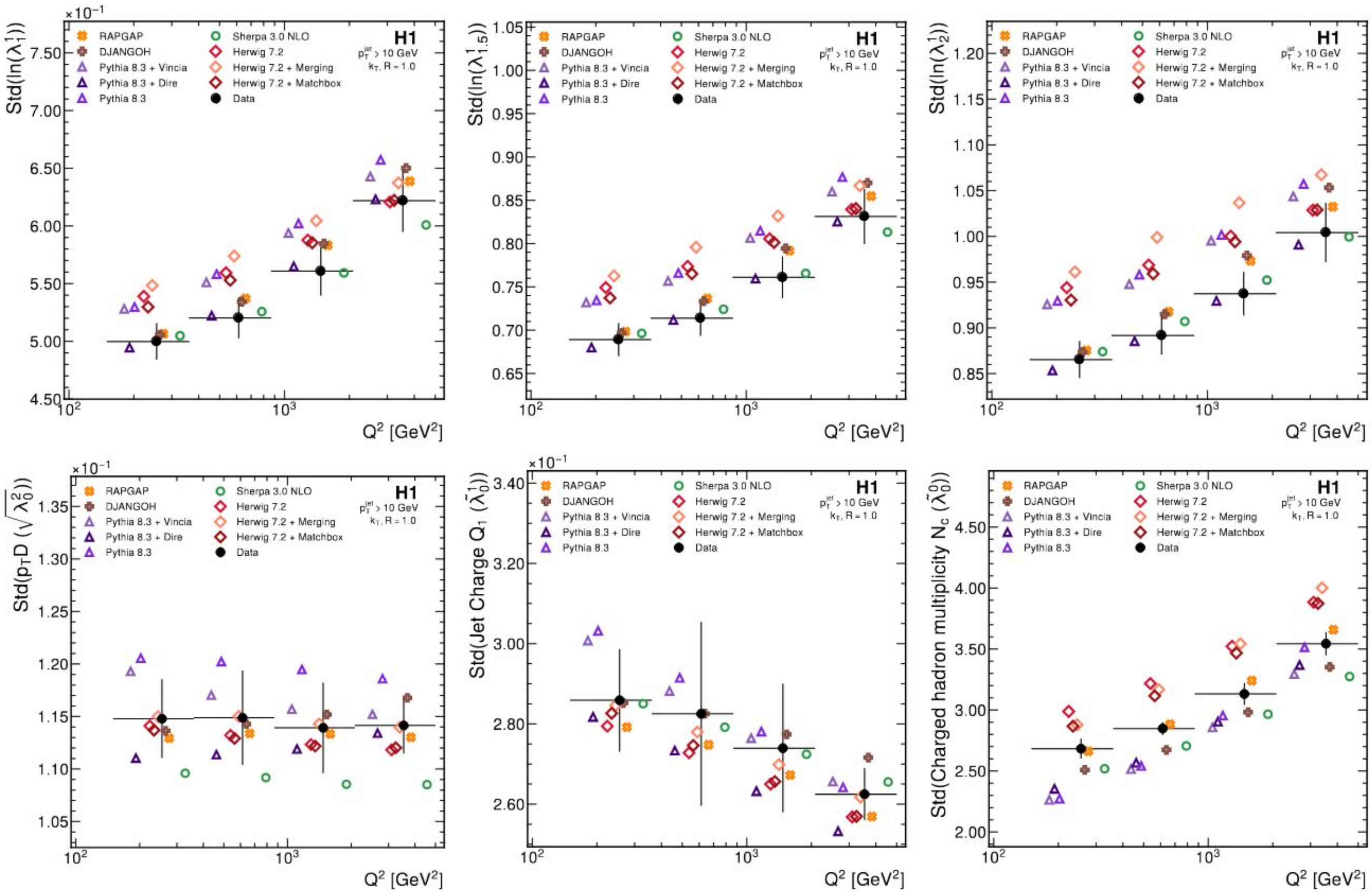


Verify the model **consistency**: start from the **Rapgap** simulation and unfold the response based on the **Djangoh** simulation

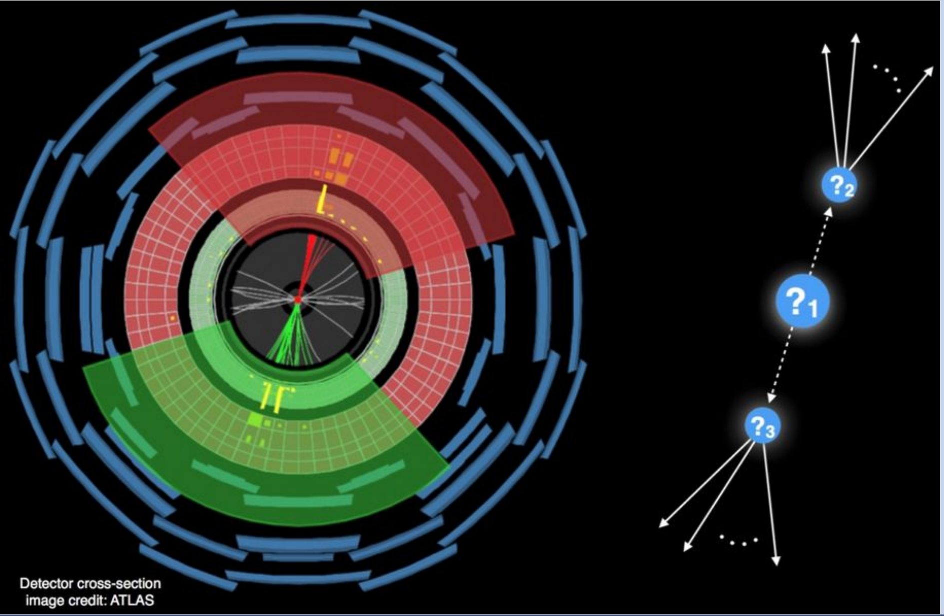
Total of **6 iterations** used to derive the main results

Multi-differential

Standard deviation of all distributions also unfolded for free

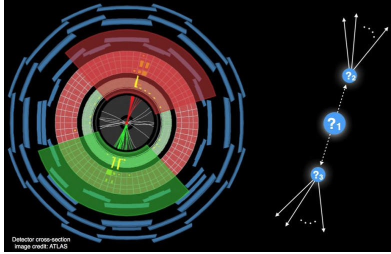


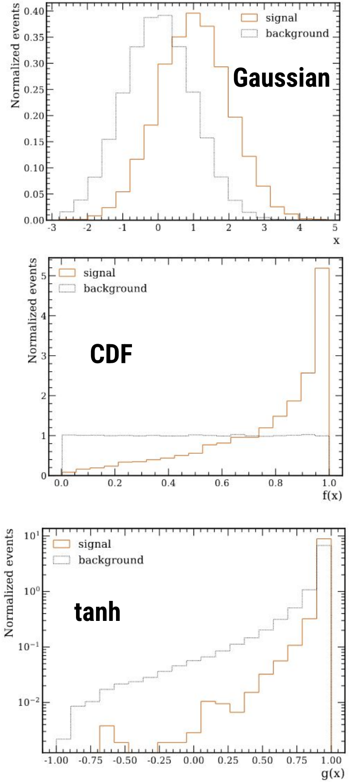
Worse general agreement between data and simulations

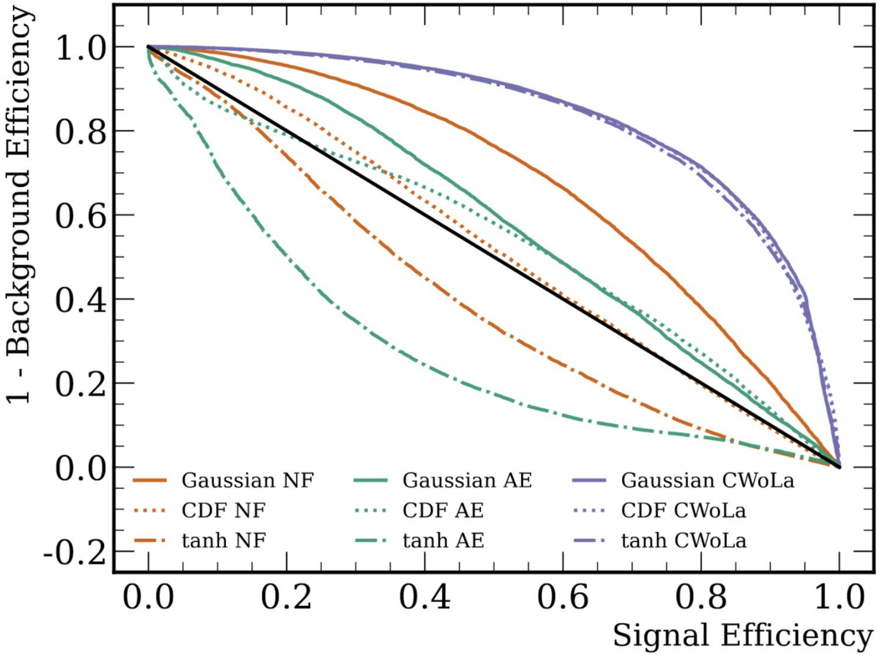


Anomaly Detection

Anomaly detection





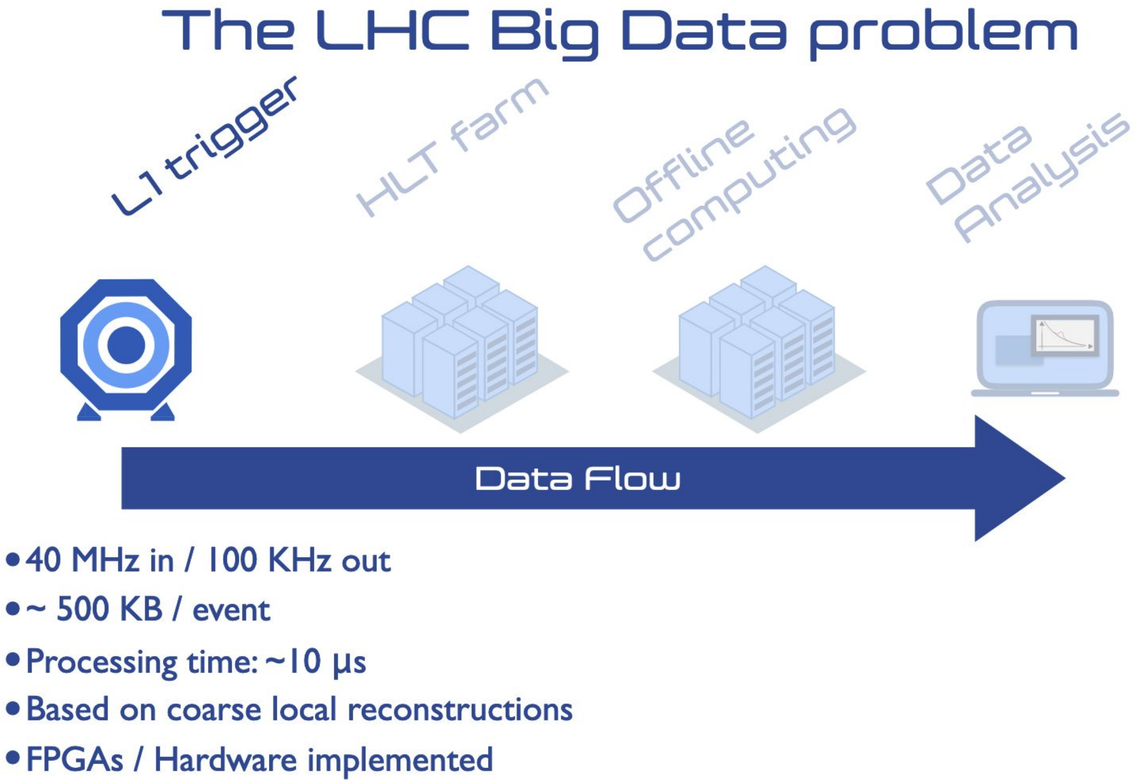


- The **set of features** used to search for anomalies can also have a big impact on the algorithm performance, as statements regarding $\mathbf{p_s(x)}$ and $\mathbf{p_b(x)}$ are not invariant under **change of coordinates**

Kasieczka, G., Mastandrea, R., **Mikuni, V.**, Nachman, B., Pettee, M., & Shih, D. (2022). *arXiv preprint arXiv:2209.06225*.

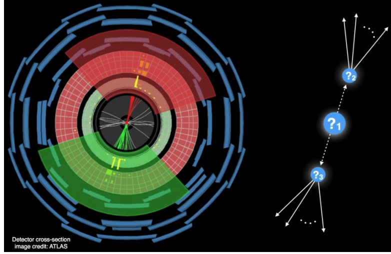
Online compatibility

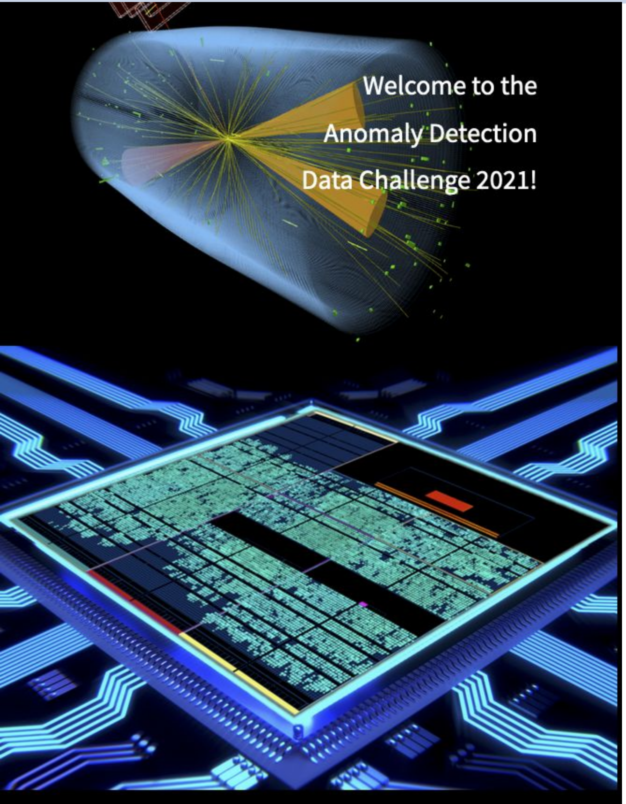
Slides from [Maurizio Pierini]



- More than **99%** of events are rejected due to **bandwidth restrictions**
- Given the algorithm's simplicity, it can also be deployed directly using modern hardware implementations such as **FPGAs**
- Possibility to identify **anomalous events** and store the information for further analysis

Anomaly detection at trigger level





- Potential to recover new physics events lost as trigger level
- Ongoing data challenge to test ideas
- CMS shows that AD triggers are feasible: **CMS-DP-2023-079**

Govorkova, Ekaterina, et al. *Scientific Data* 9.1 (2022): 118.
Govorkova, Ekaterina, et al. *Nature Machine Intelligence* 4.2 (2022): 154-161.