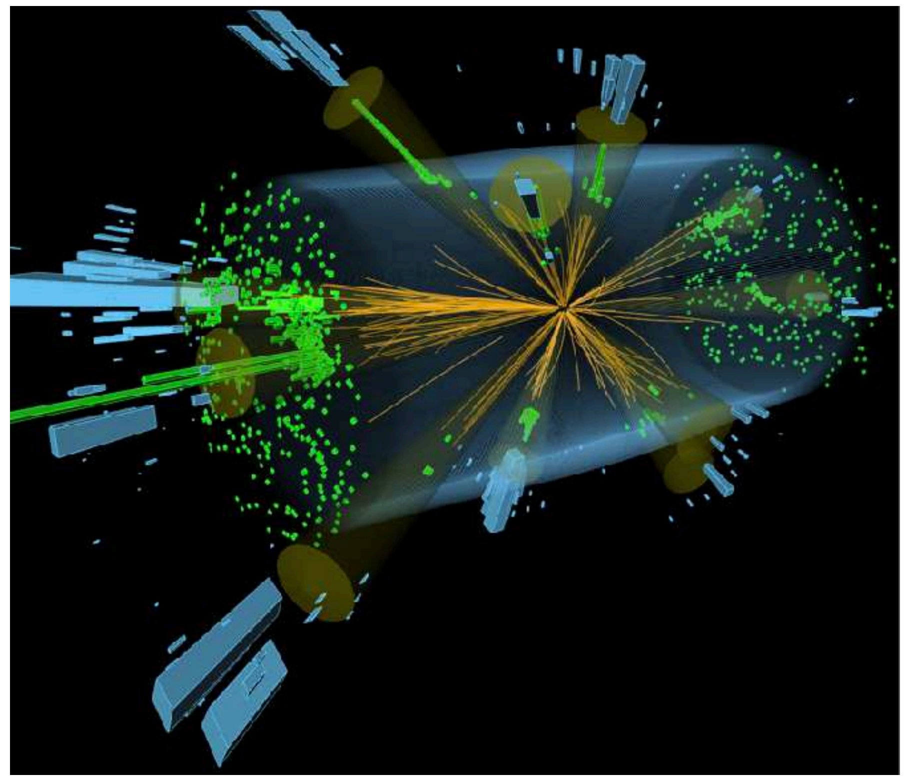
Optimizing Particle Physics with Machine Learning

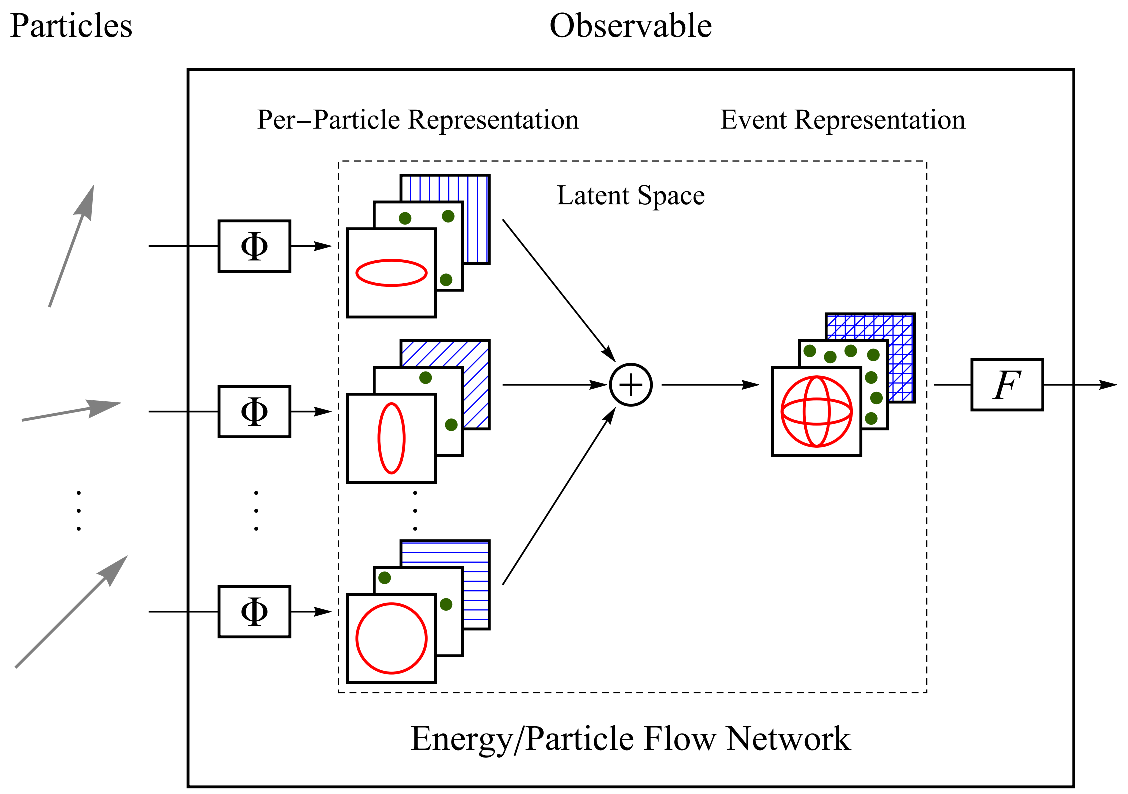
Patrick T. Komiske III
Massachusetts Institute of Technology
Center for Theoretical Physics
NSF AI Institute for Artificial Intelligence and Fundamental Interactions

MIT Lincoln Laboratory

Groups 41 & 89 Joint Seminar

May 12, 2021

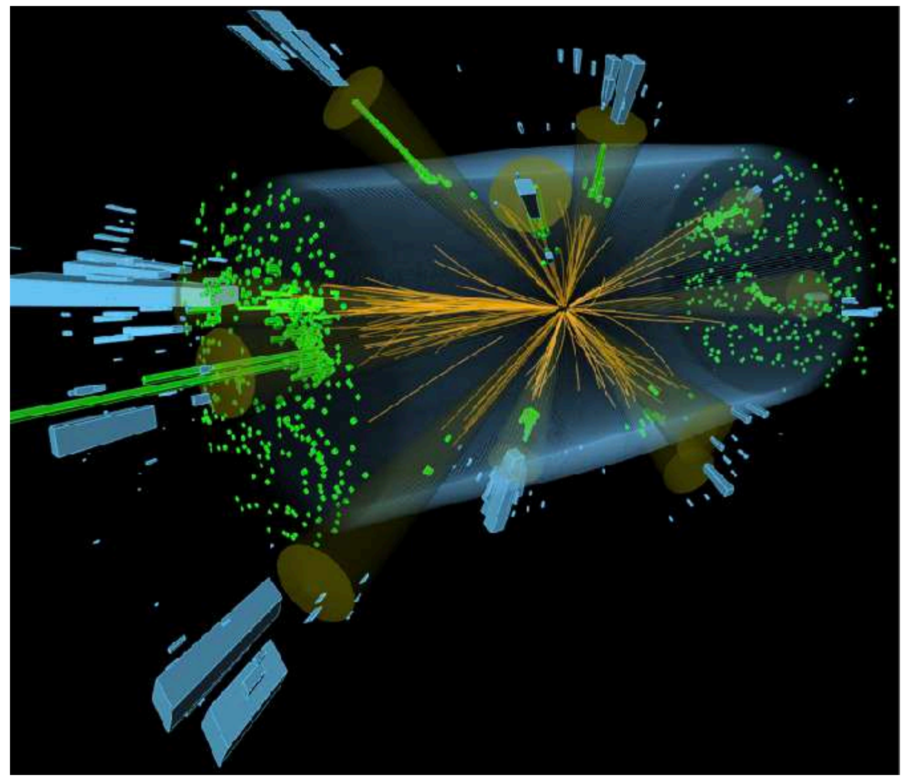




Particle Physics Fundamentals

Architectures for Colliders

Statistical Deconvolution



Particle Physics Fundamentals

Why do we collide particles at ultra-high energies?

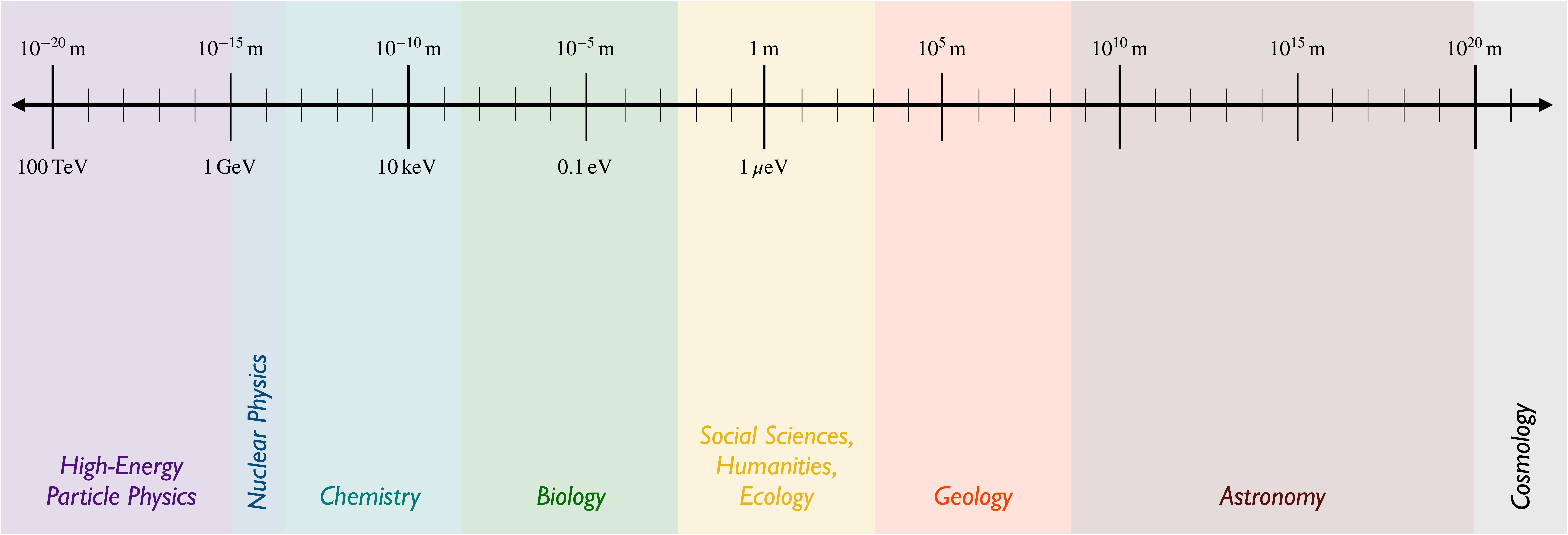
What are some outstanding challenges that ML can help address?

Architectures for Colliders

Statistical Deconvolution

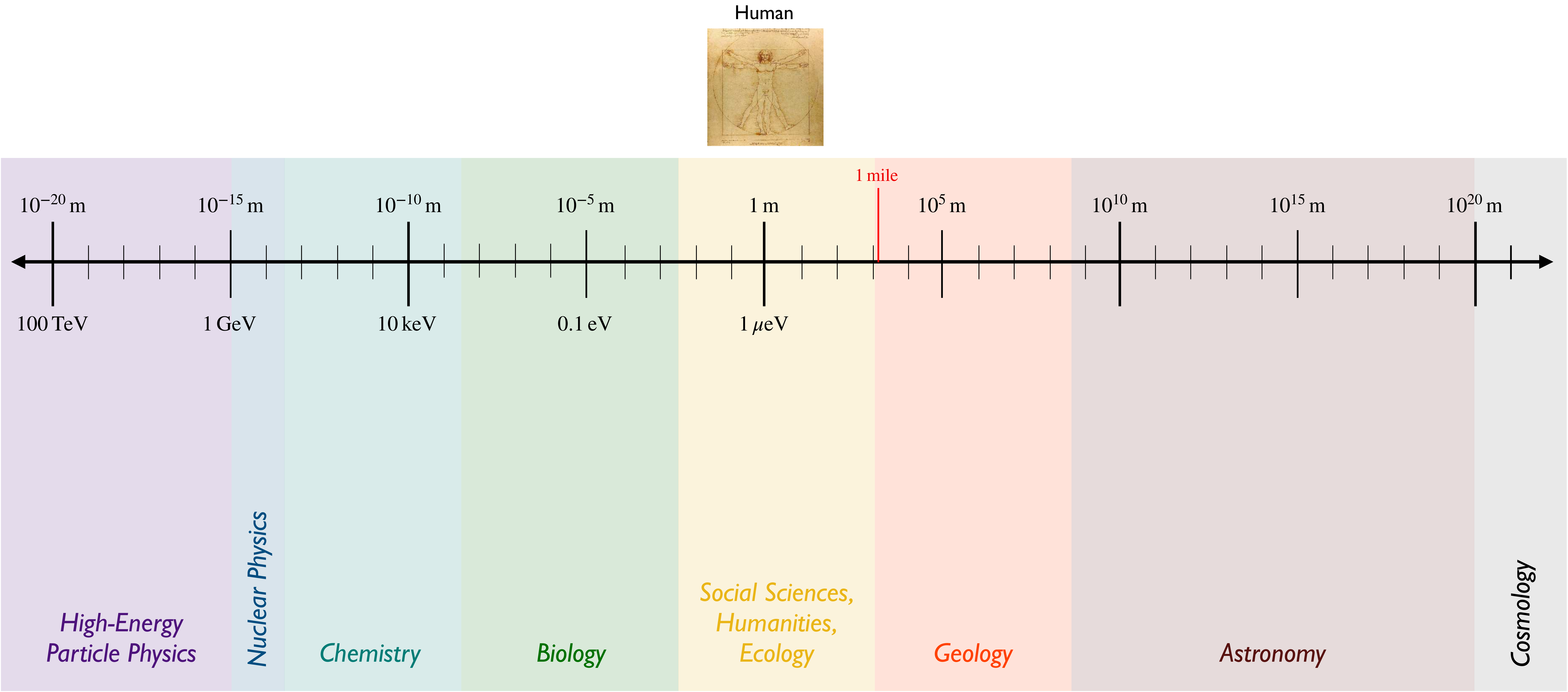
Length/Energy Scales in Nature

Einstein-Planck Equation: $E = \frac{hc}{\lambda}$



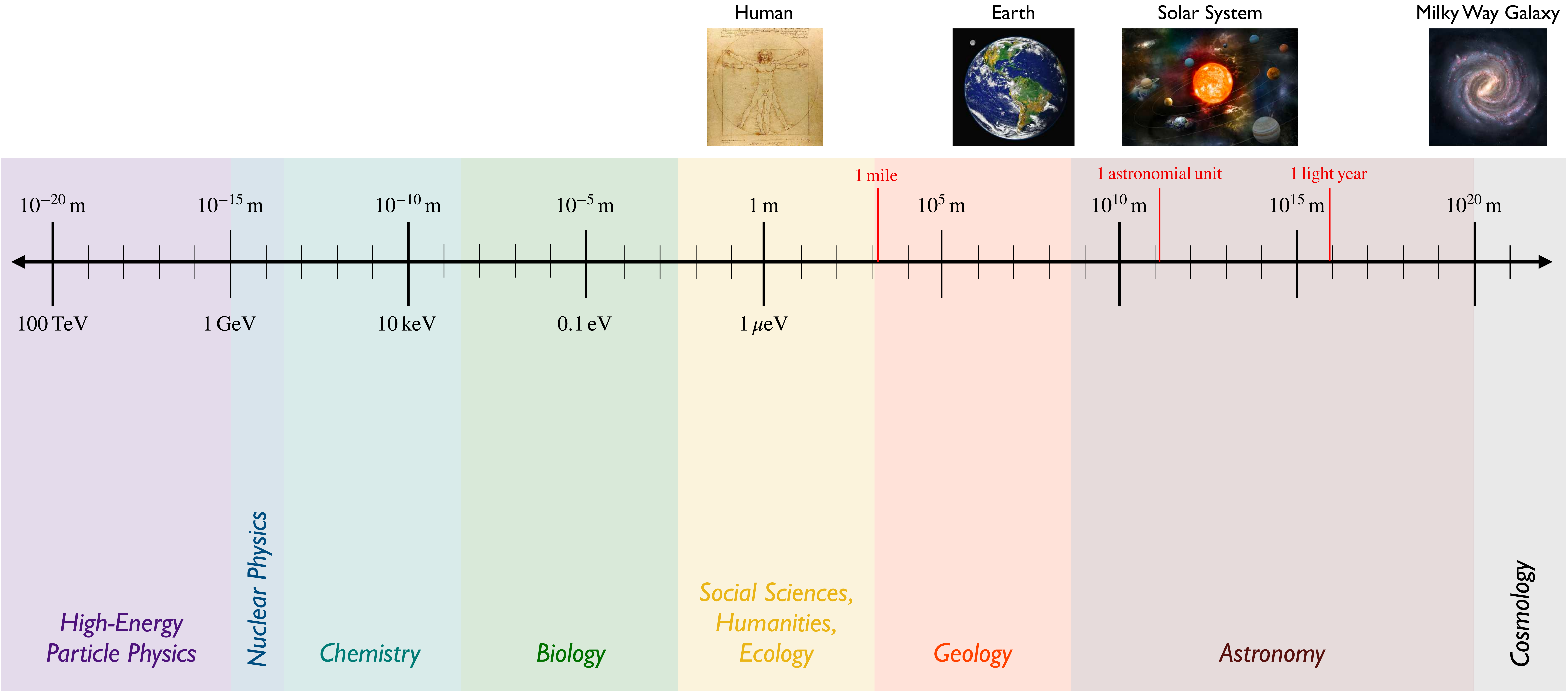
Length/Energy Scales in Nature

Einstein-Planck Equation: $E = \frac{hc}{\lambda}$



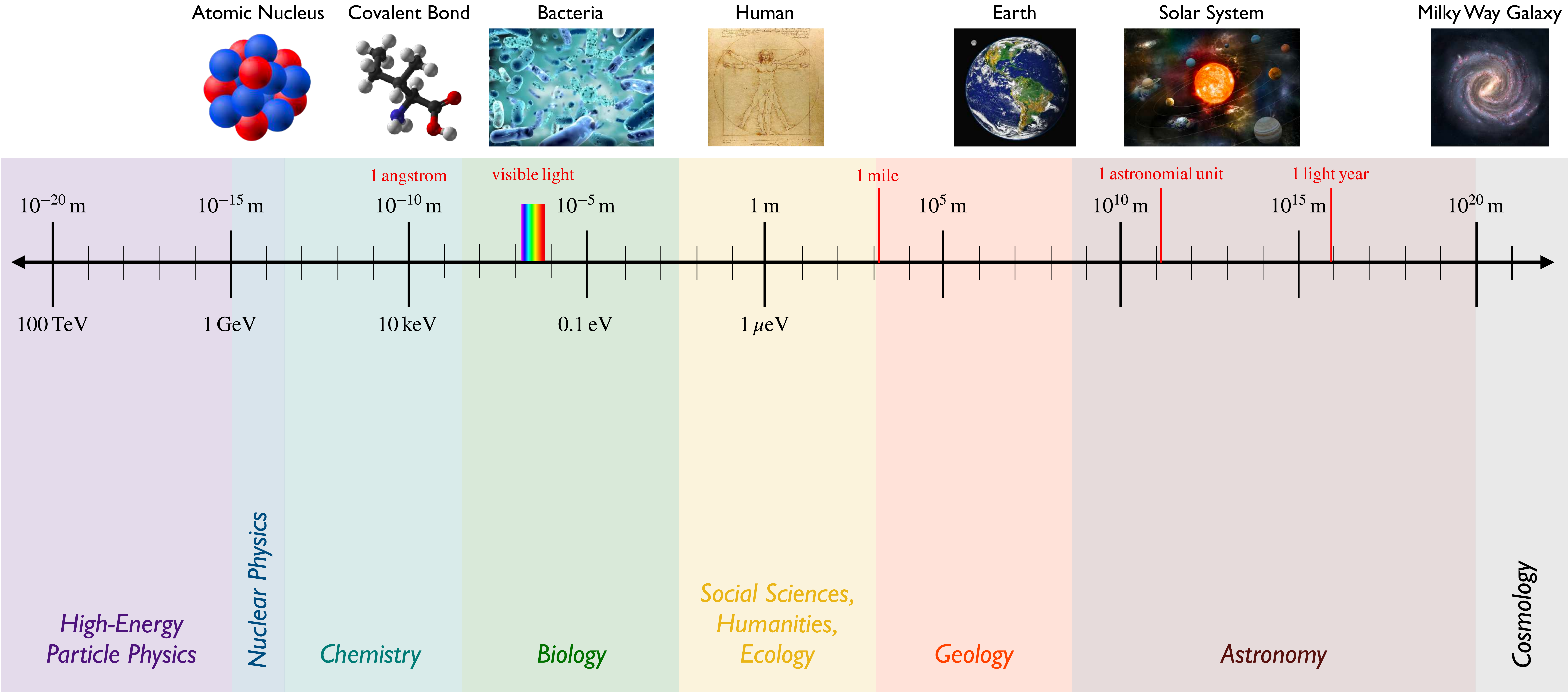
Length/Energy Scales in Nature

Einstein-Planck Equation: $E = \frac{hc}{\lambda}$



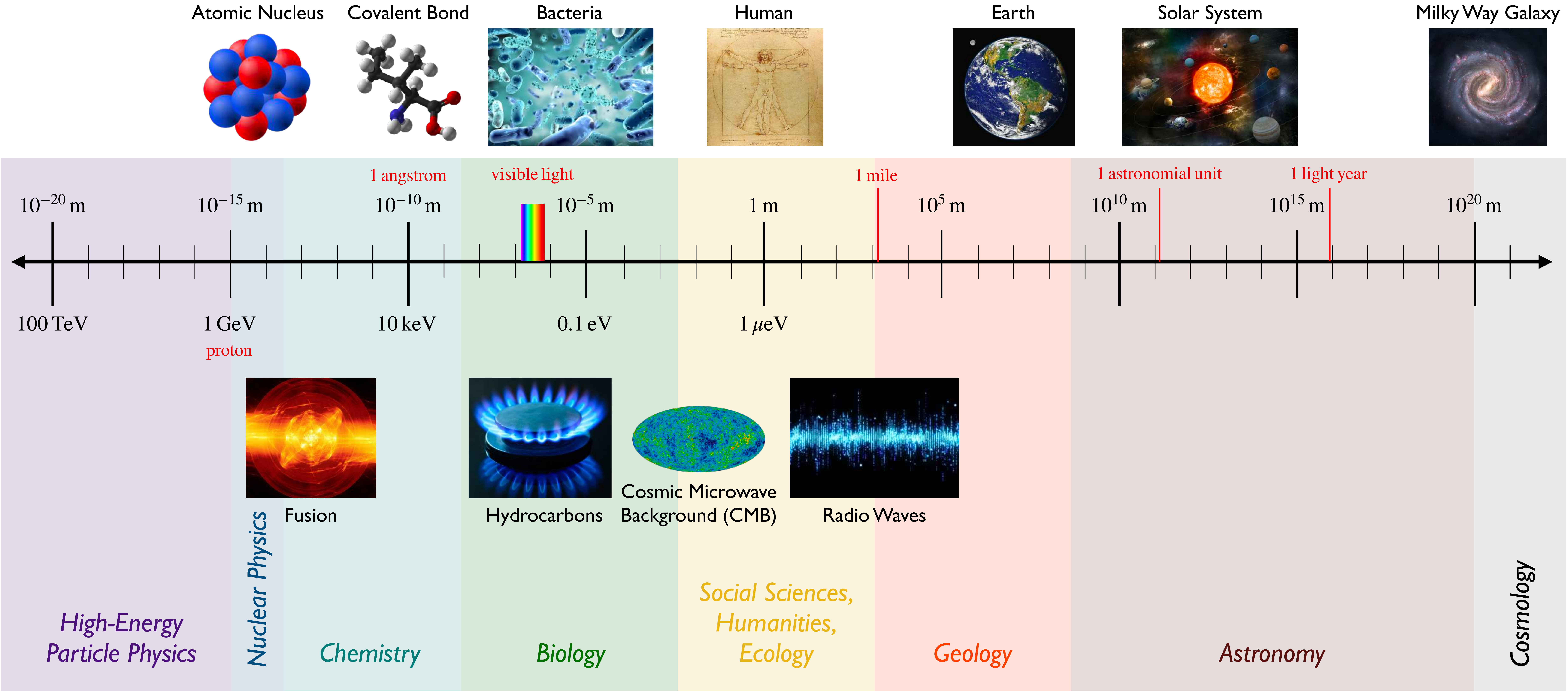
Length/Energy Scales in Nature

Einstein-Planck Equation: $E = \frac{hc}{\lambda}$



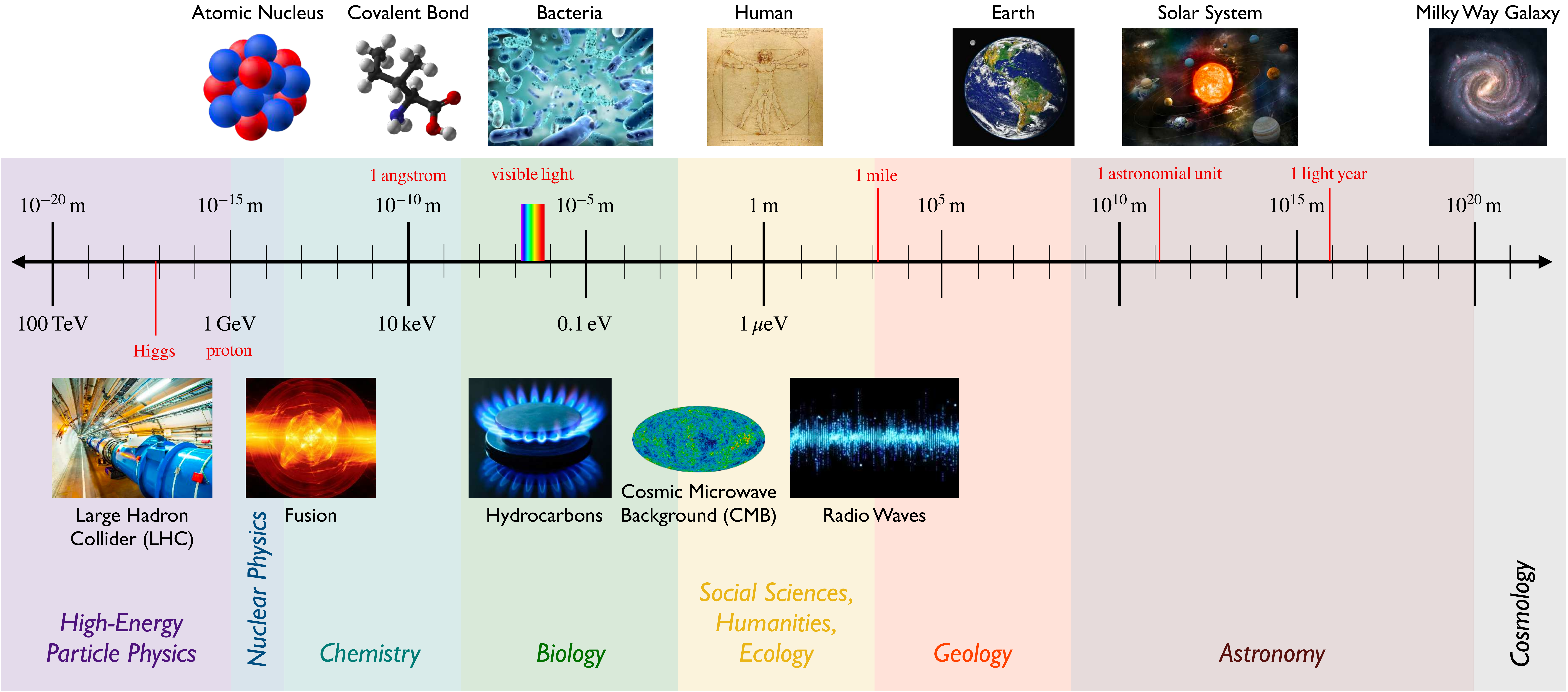
Length/Energy Scales in Nature

Einstein-Planck Equation: $E = \frac{hc}{\lambda}$



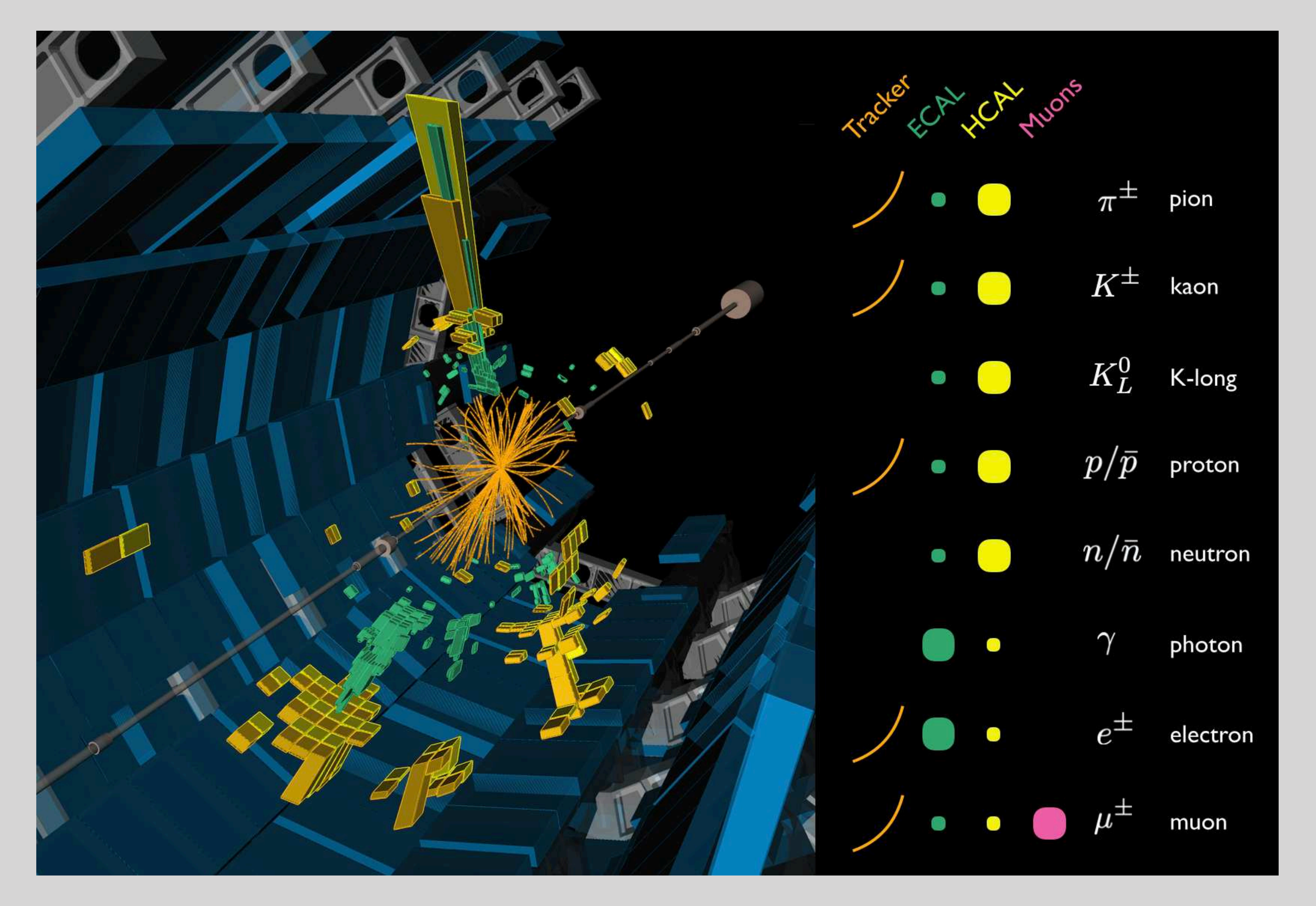
Length/Energy Scales in Nature

Einstein-Planck Equation: $E = \frac{hc}{\lambda}$



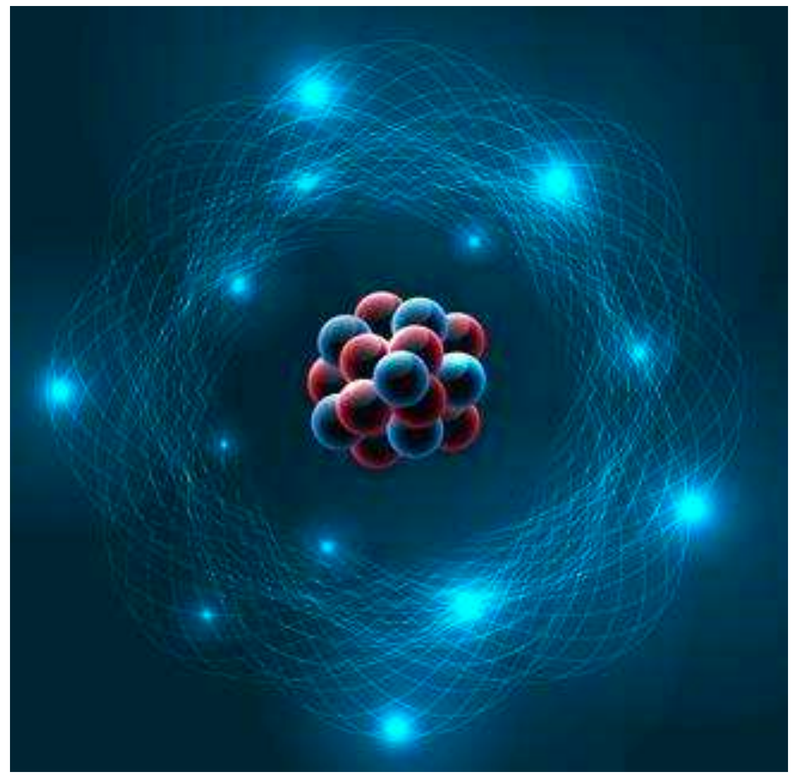
Events at the LHC

High-energy proton-proton collisions produce particles with energy, direction, charge, and flavor

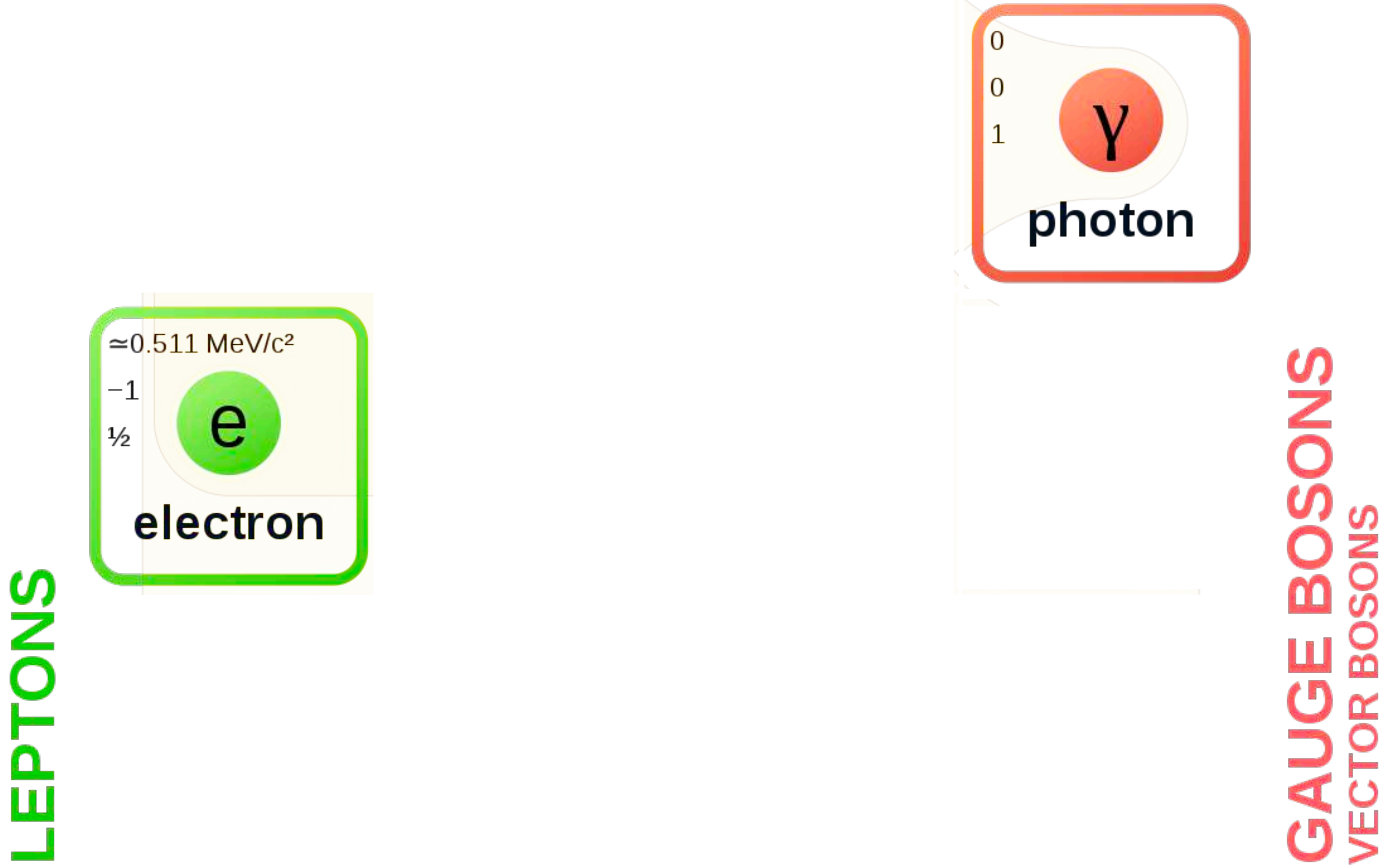


Standard Model of Particle Physics at "Low"-Energies

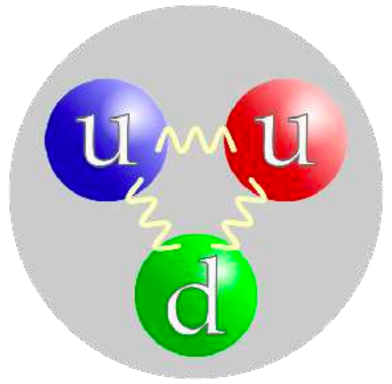
Standard Model of Particle Physics at "Low"-Energies

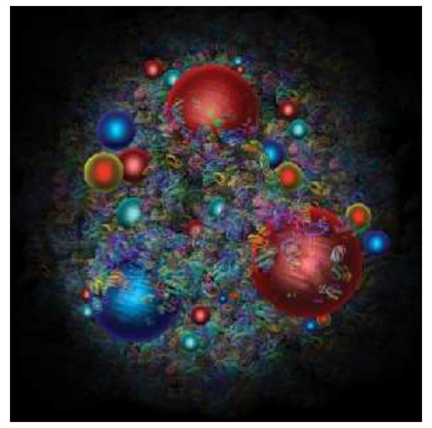


Atomic ingredients
– nucleus of protons and neutrons
– electron cloud around nucleus
– bound by electromagnetism

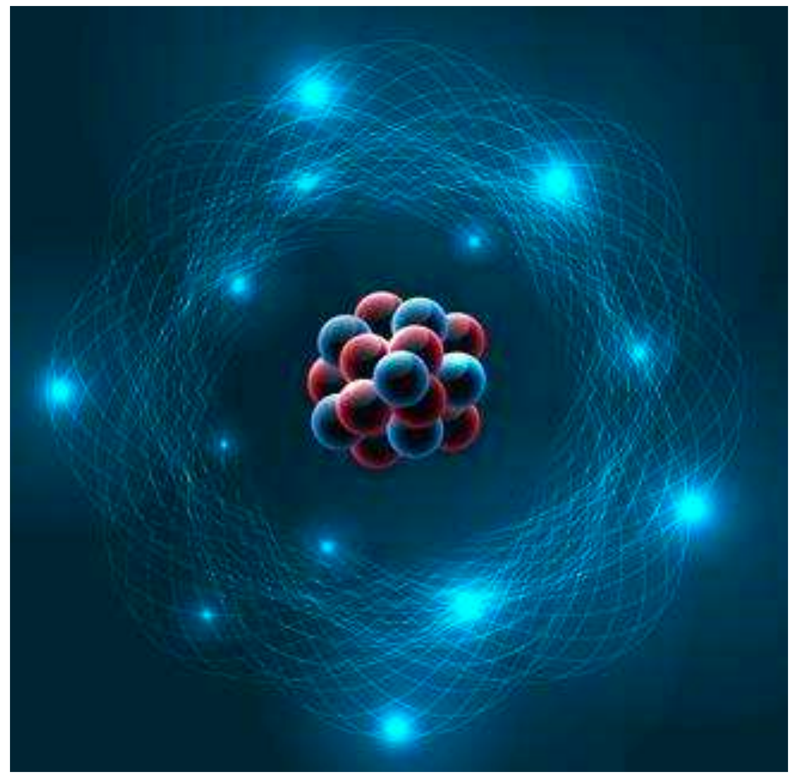


Standard Model of Particle Physics at "Low"-Energies

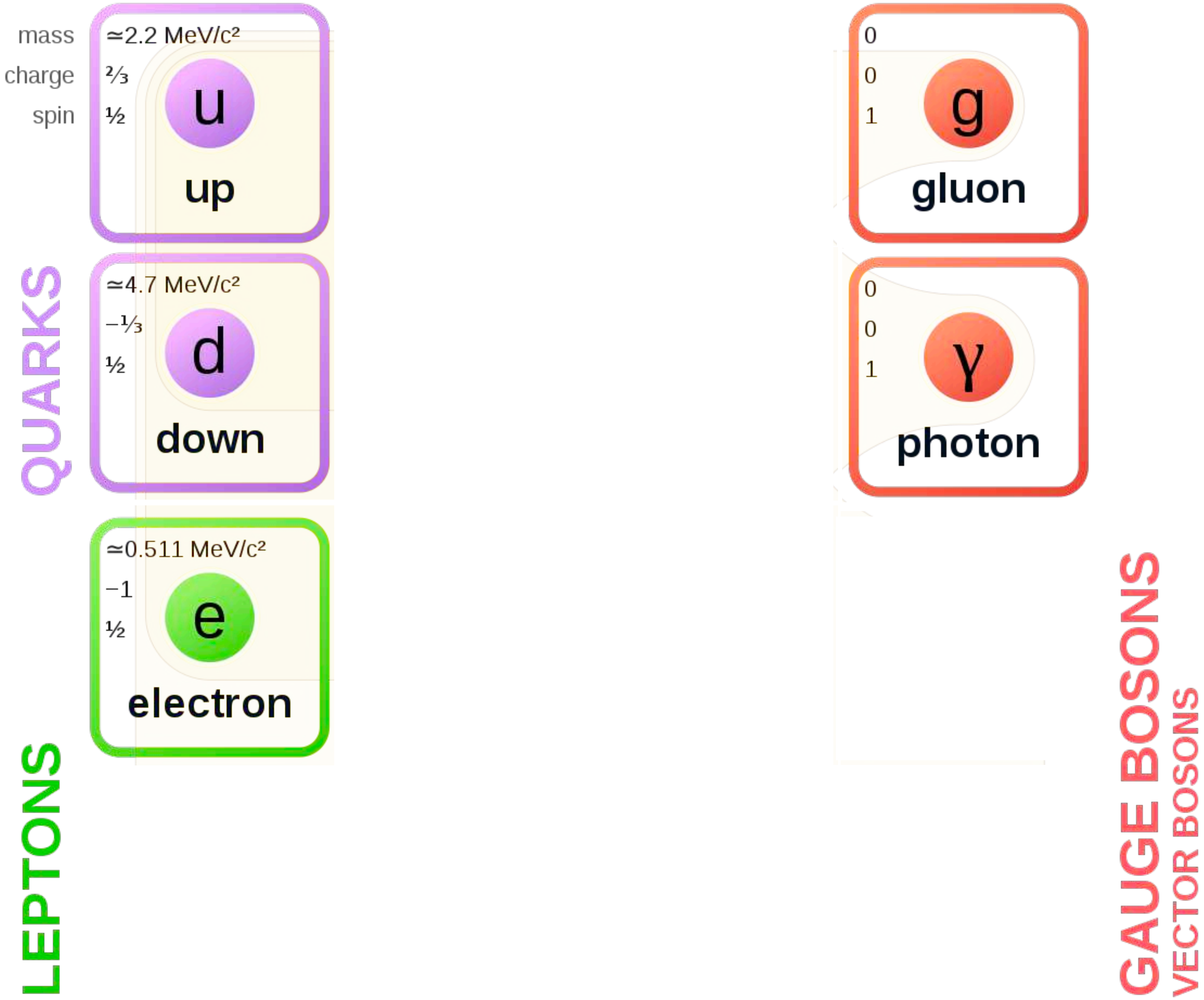




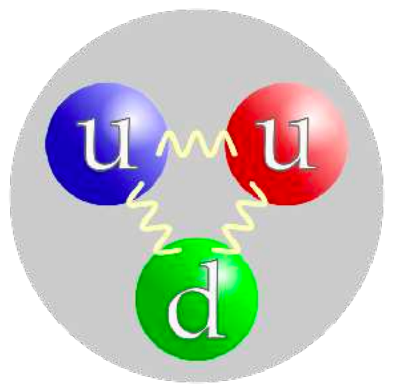
proton (uud)
neutron (udd)

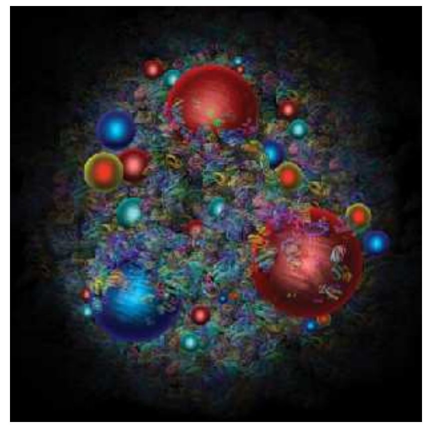


Atomic ingredients
– nucleus of protons and neutrons
– electron cloud around nucleus
– bound by electromagnetism

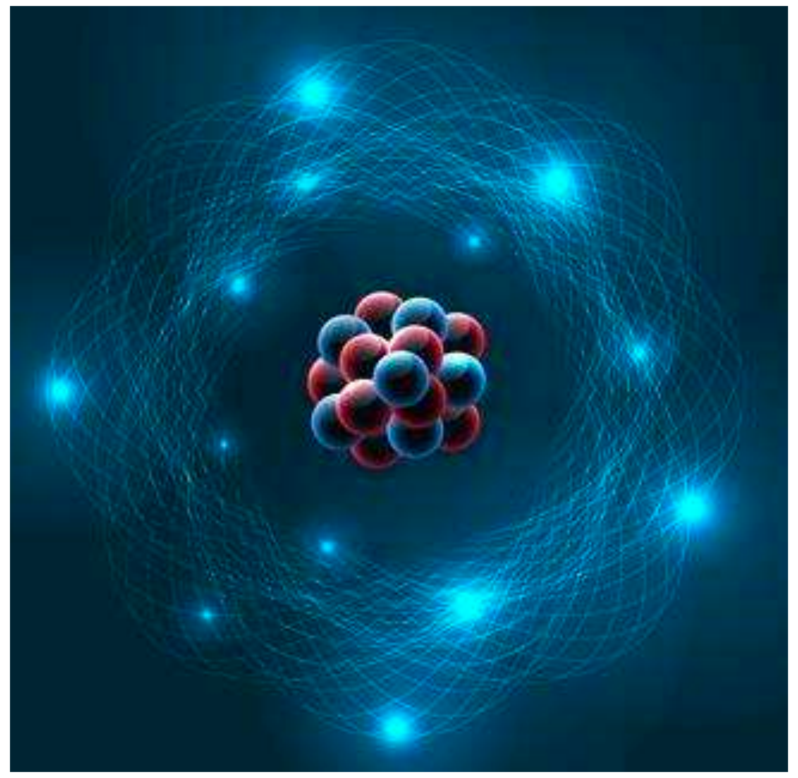


Standard Model of Particle Physics at “Low”-Energies



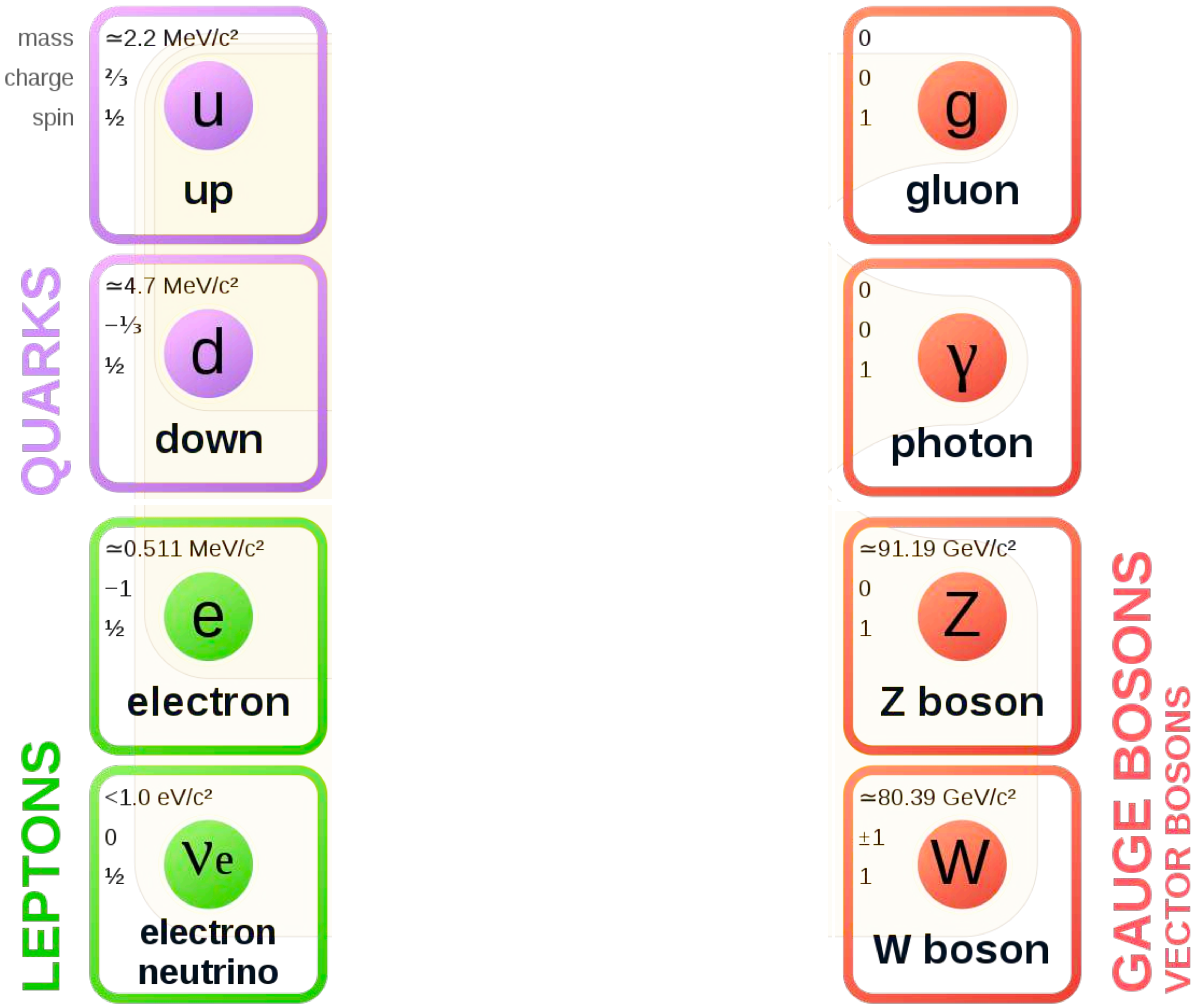


proton (uud)
neutron (udd)



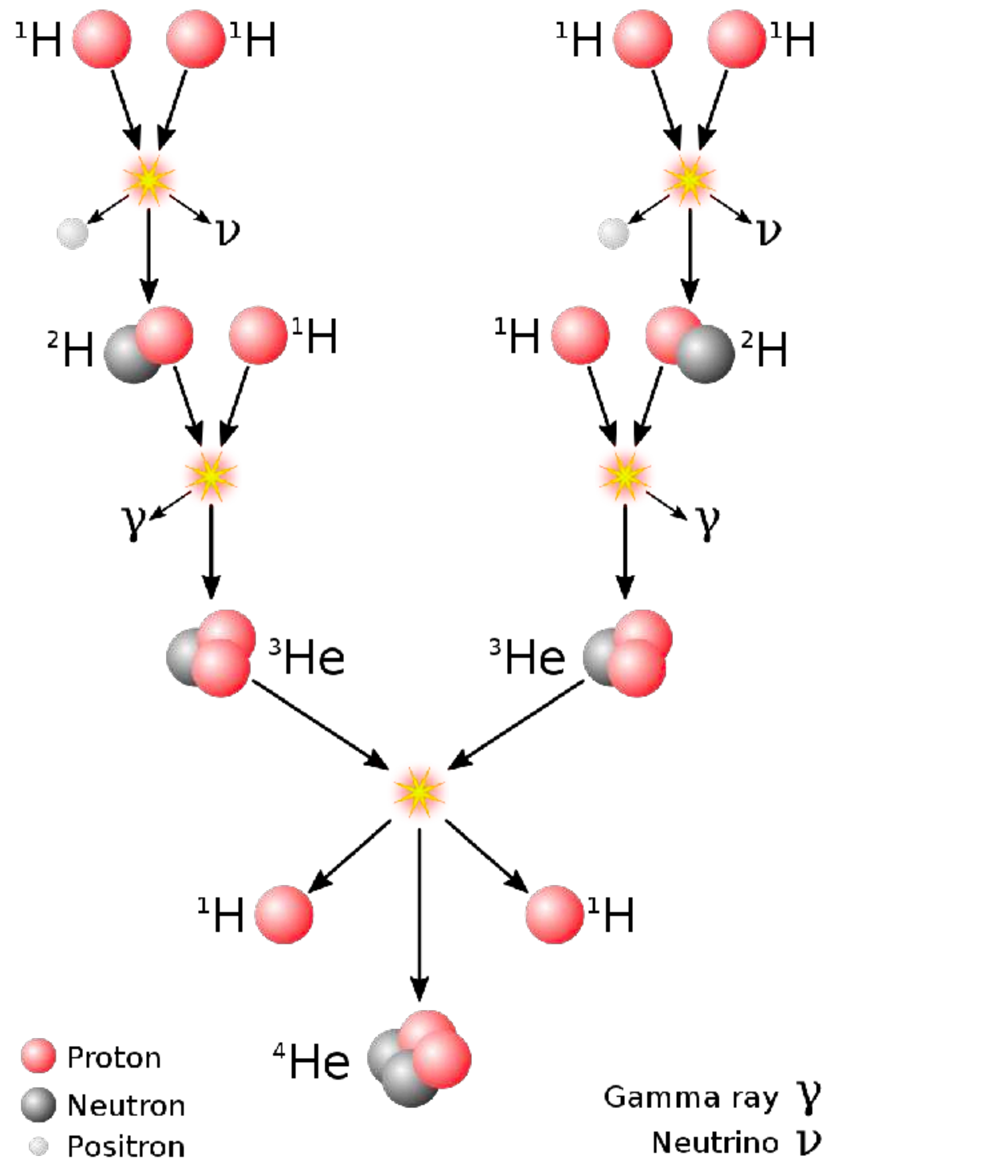
Atomic ingredients

– nucleus of protons and neutrons
– electron cloud around nucleus
– bound by electromagnetism

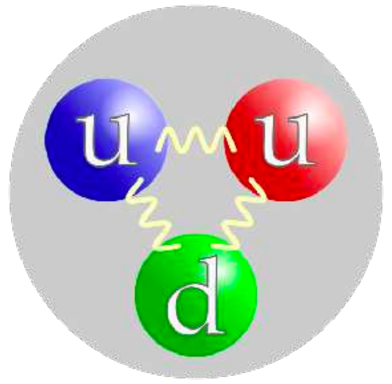


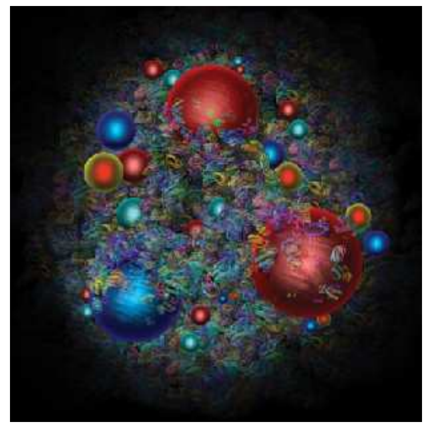
Solar nuclear fusion

– protons first fuse into deuterium via weak force (a.k.a. W boson)
– ~9 billion years for average proton to fuse

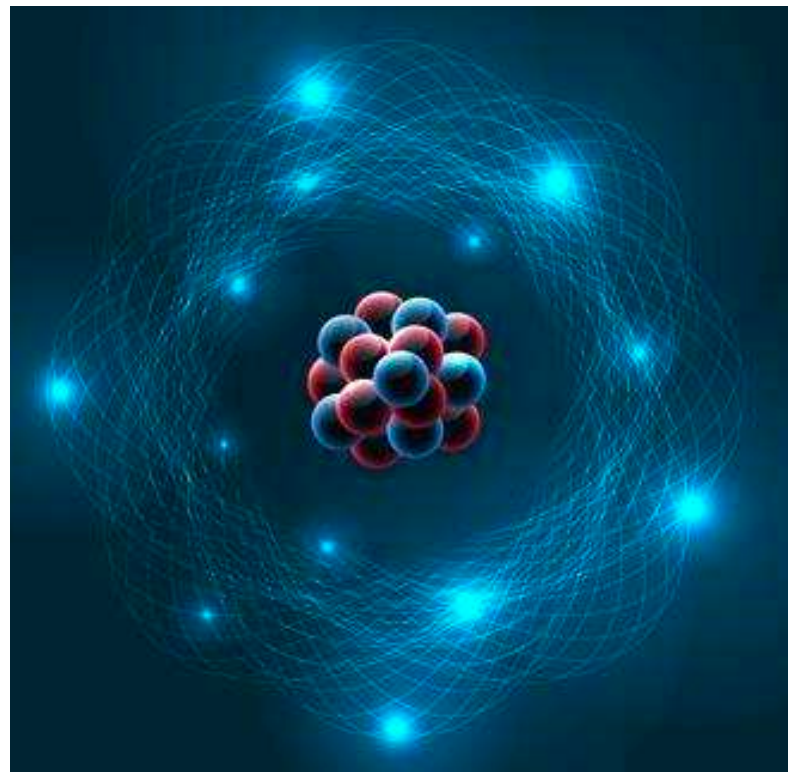


Standard Model of Particle Physics at "Low"-Energies

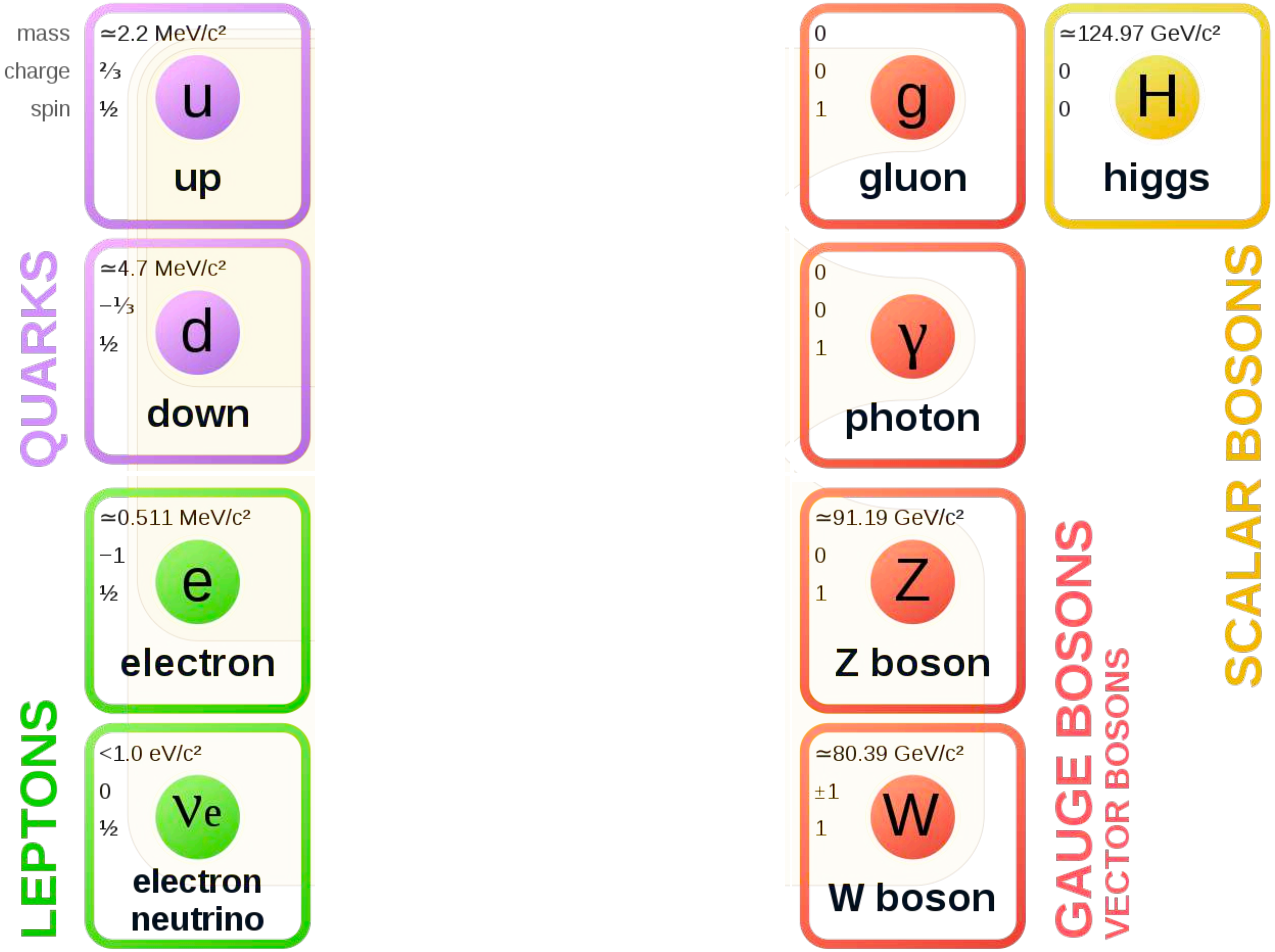


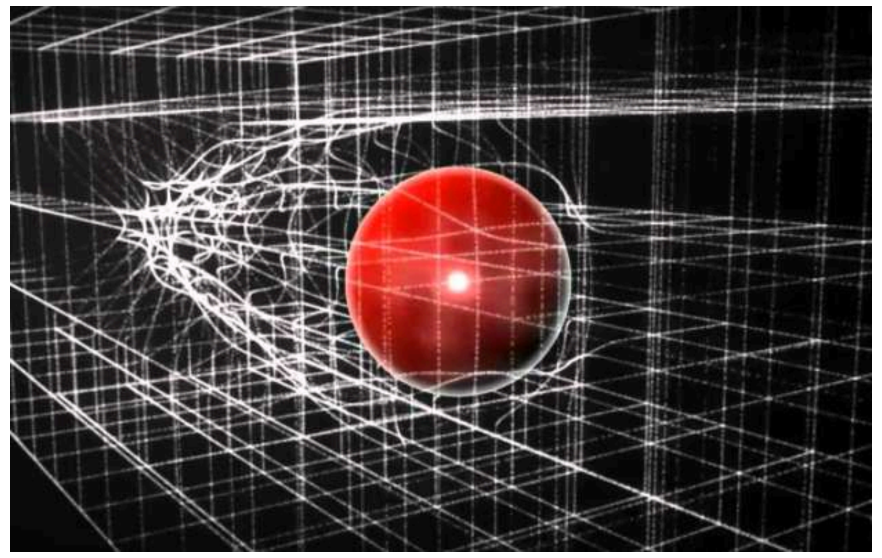


proton (uud)
neutron (udd)



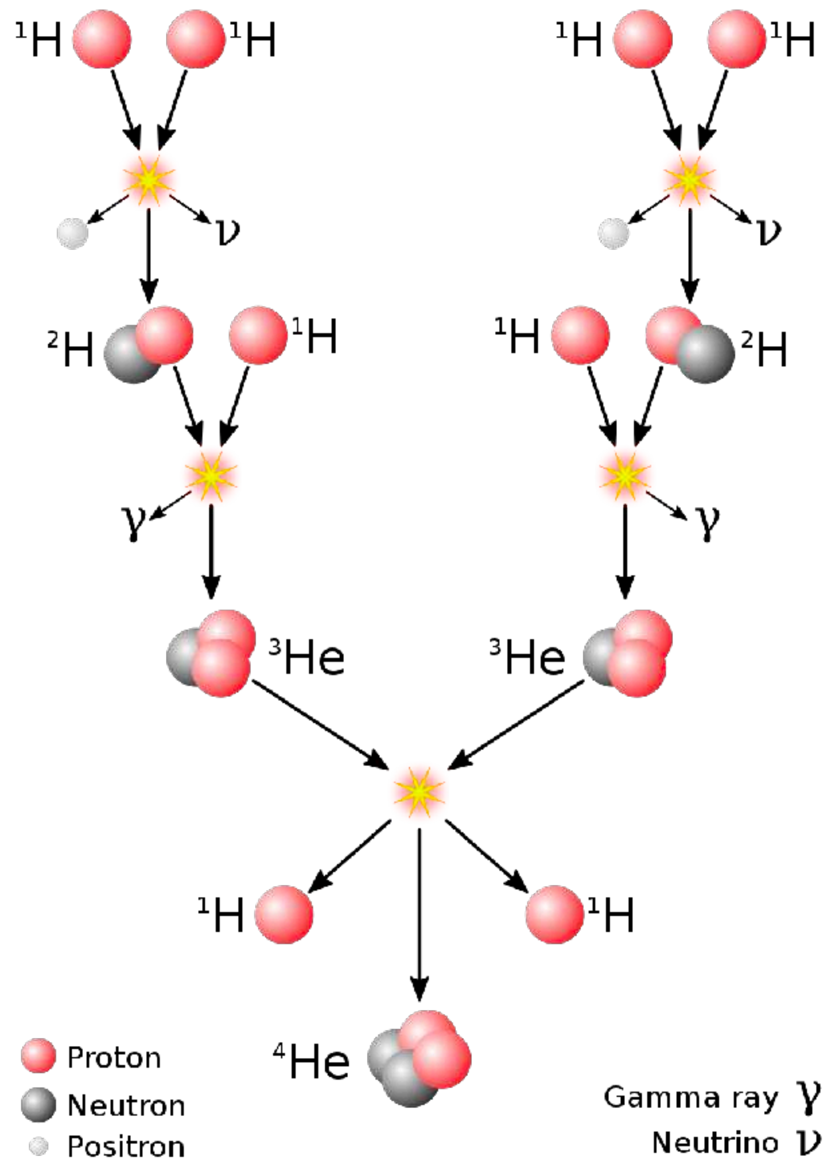
Atomic ingredients
– nucleus of protons and neutrons
– electron cloud around nucleus
– bound by electromagnetism



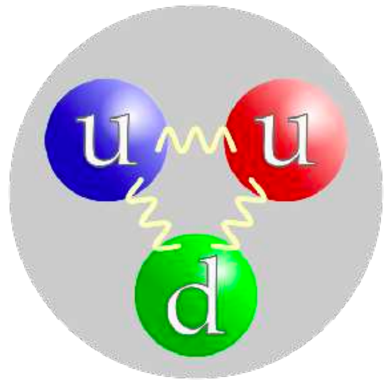


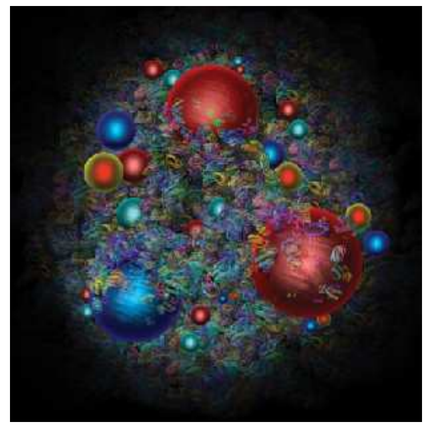
Particle mass comes from interaction with Higgs field

Solar nuclear fusion
– protons first fuse into deuterium via weak force (a.k.a. W boson)
– ~9 billion years for average proton to fuse

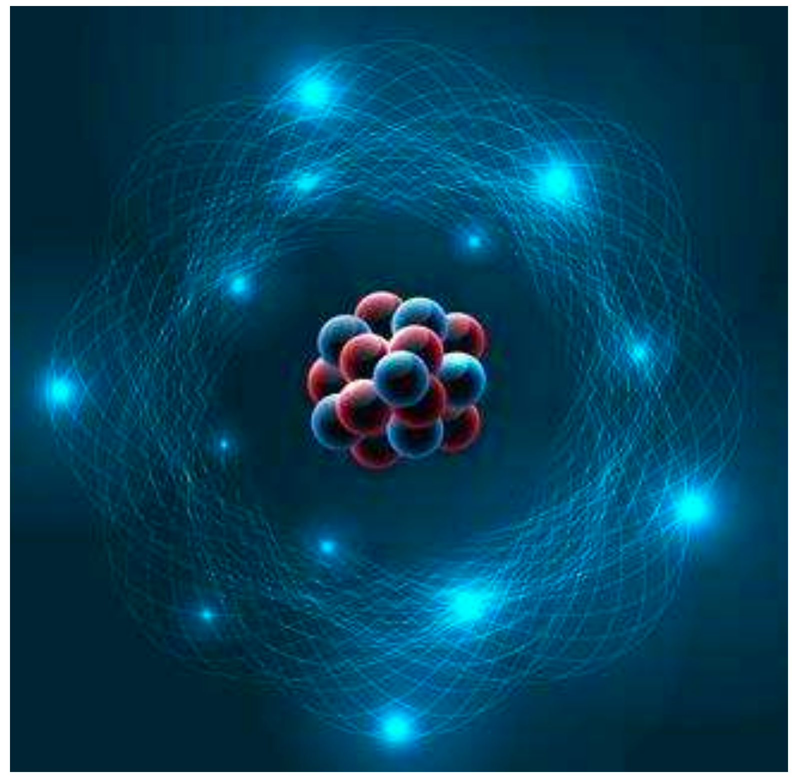


Standard Model of Particle Physics at "Low"-Energies



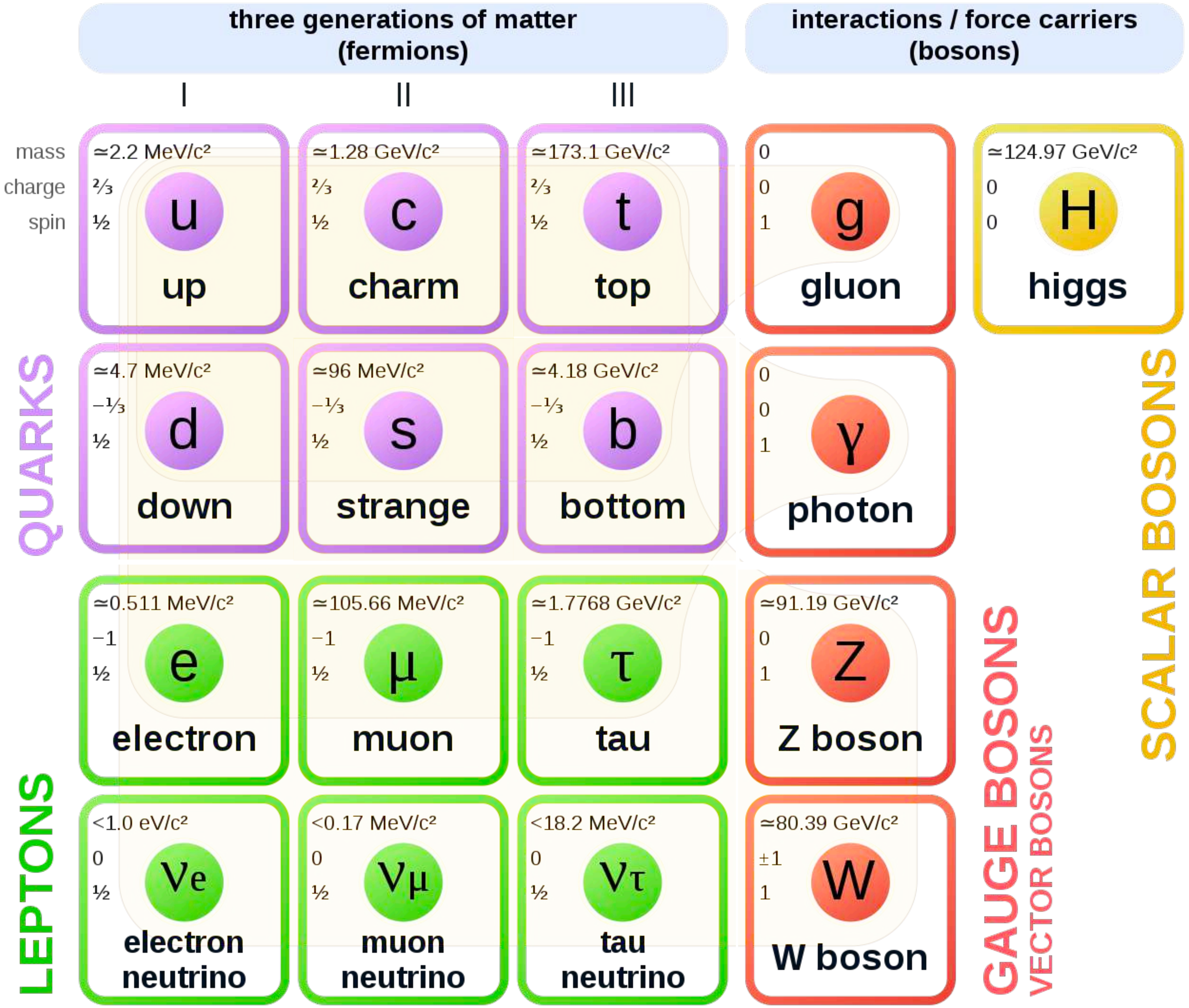


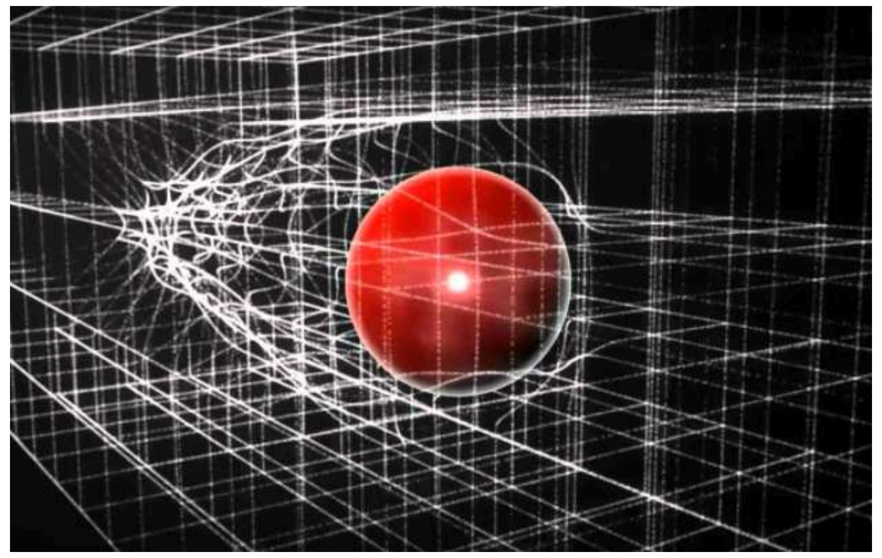
proton (uud)
neutron (udd)



Atomic ingredients

– nucleus of protons and neutrons
– electron cloud around nucleus
– bound by electromagnetism

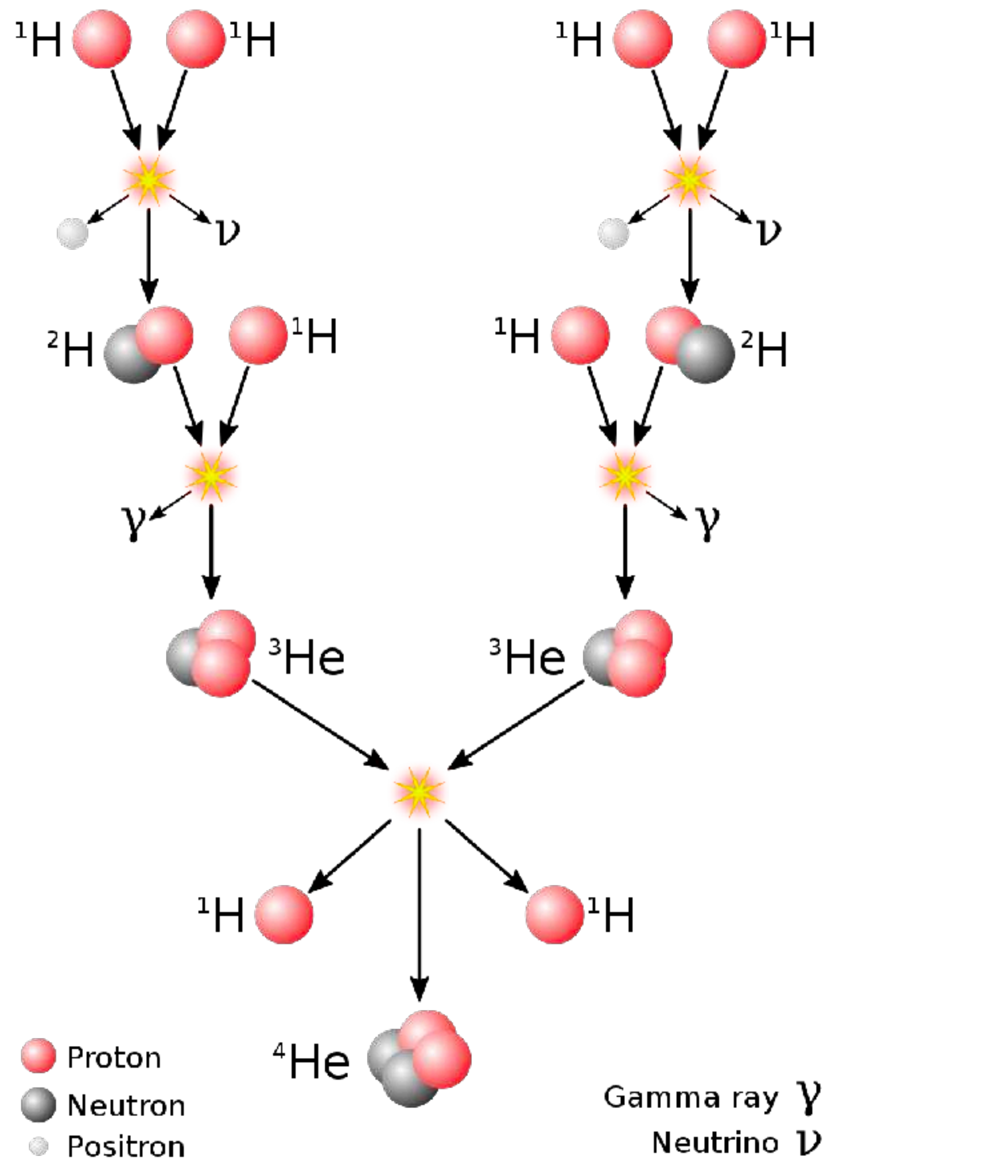




Particle mass comes from interaction with Higgs field

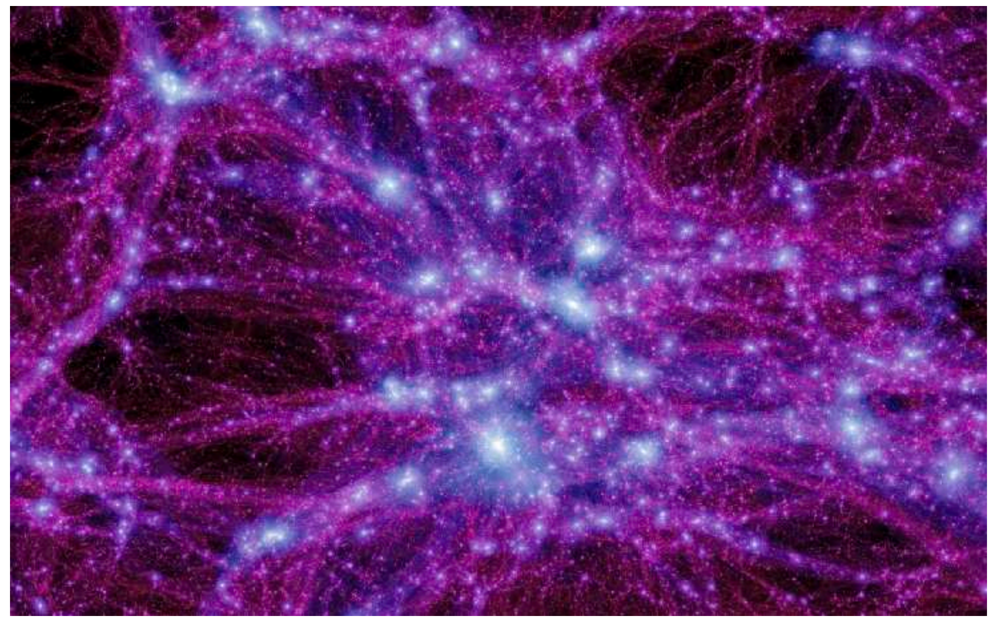
Solar nuclear fusion

– protons first fuse into deuterium via weak force (a.k.a. W boson)
– ~9 billion years for average proton to fuse



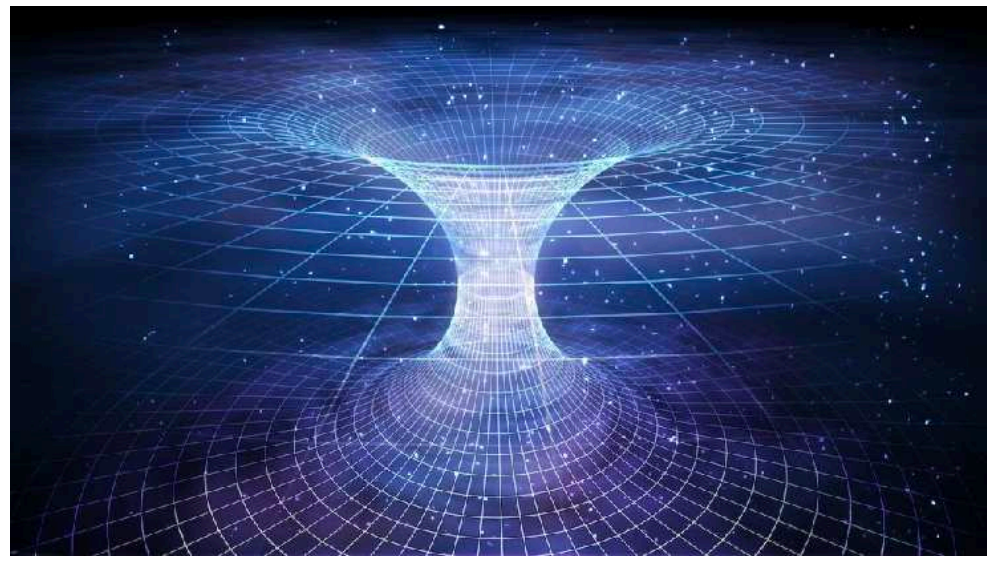
Standard Model of Particle Physics – Unanswered Questions

Dark matter?

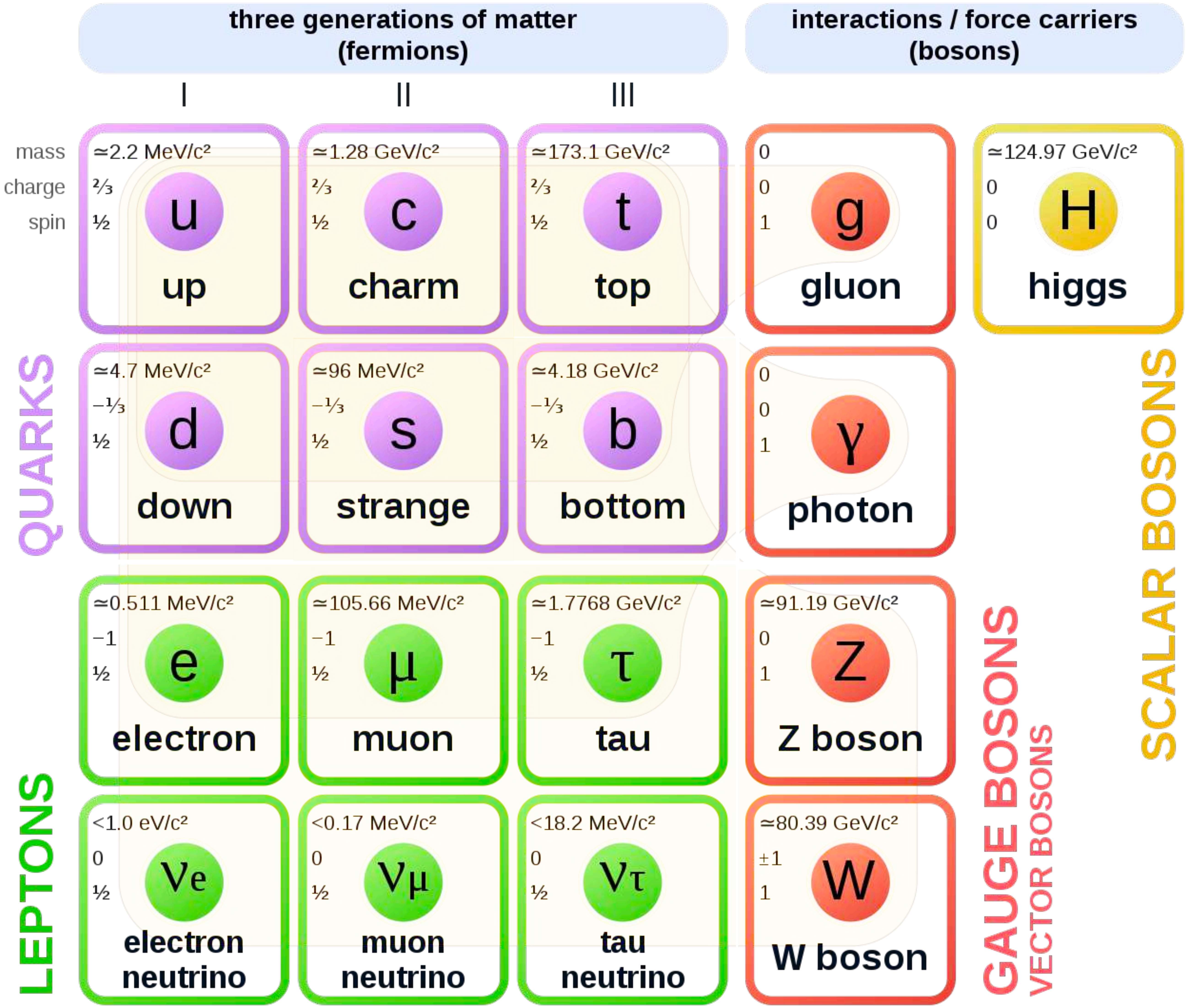


≈25% of the energy of the universe

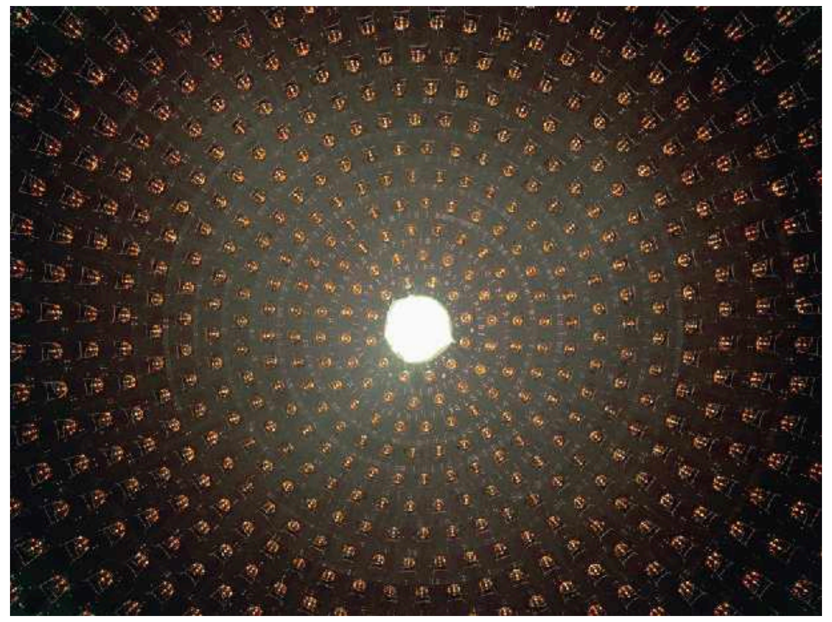
Acceleration of universe?



≈68% of the universe is “dark energy”

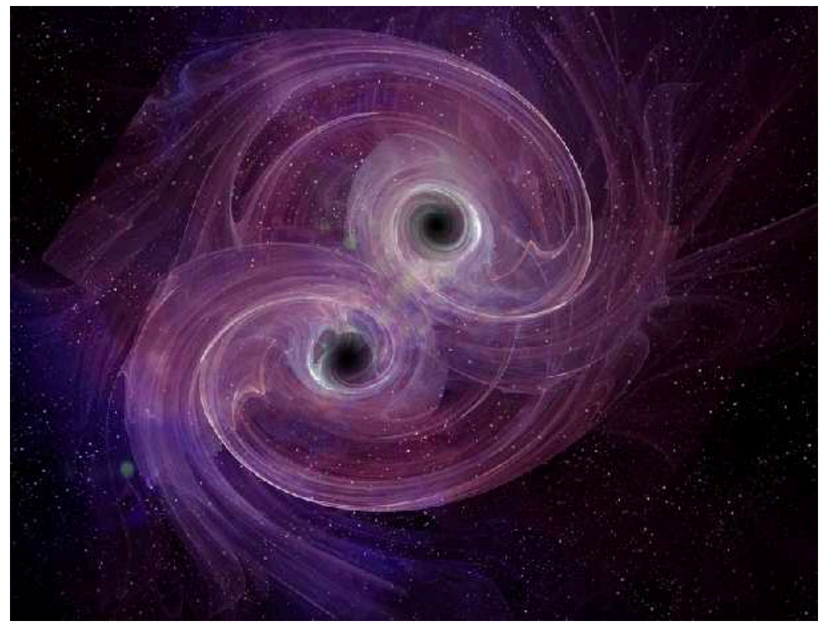


Matter/anti-matter asymmetry?



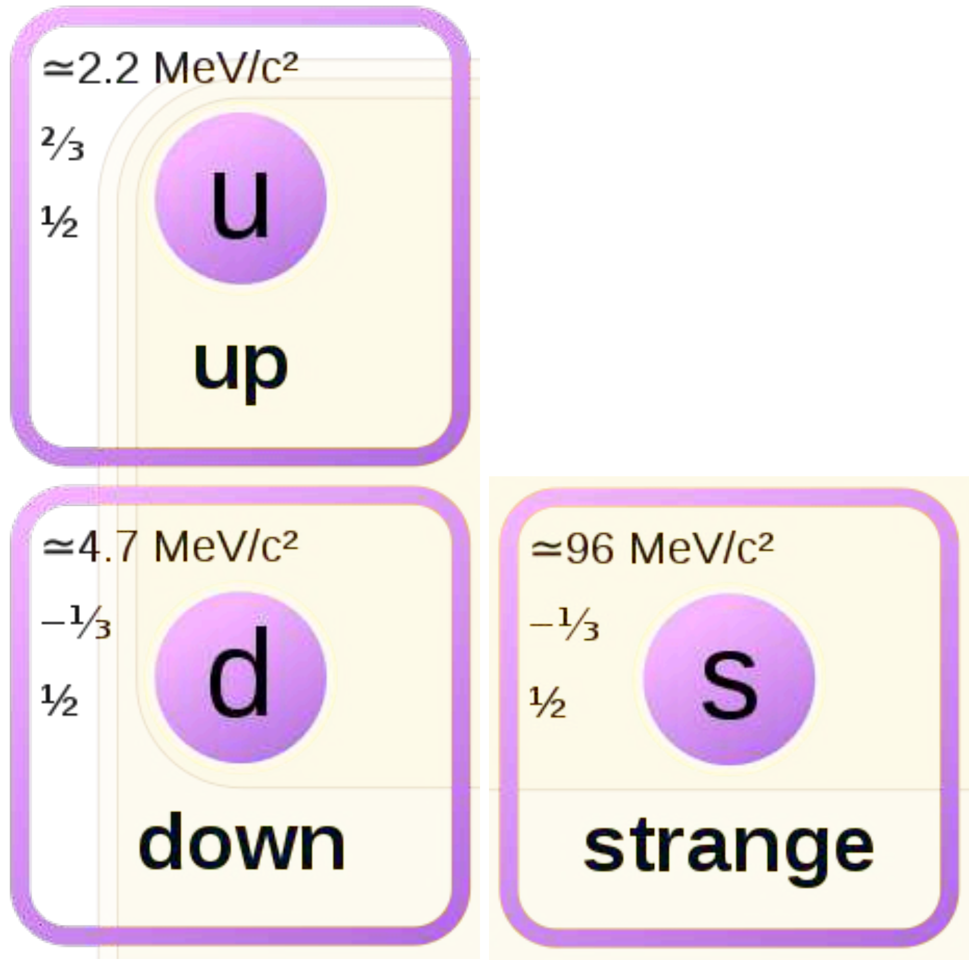
MiniBooNE neutrino detector

What about gravity?



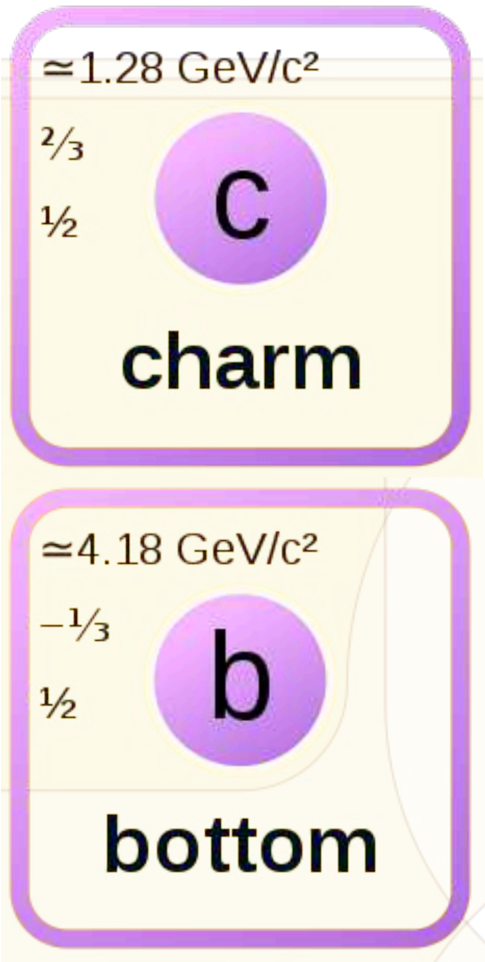
Standard Model of Particle Physics at High-Energies – as Jets

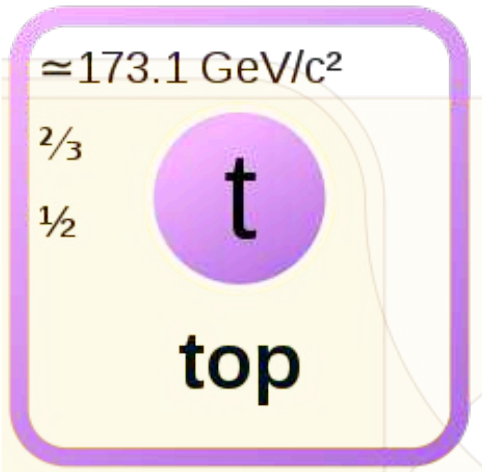
Light quarks

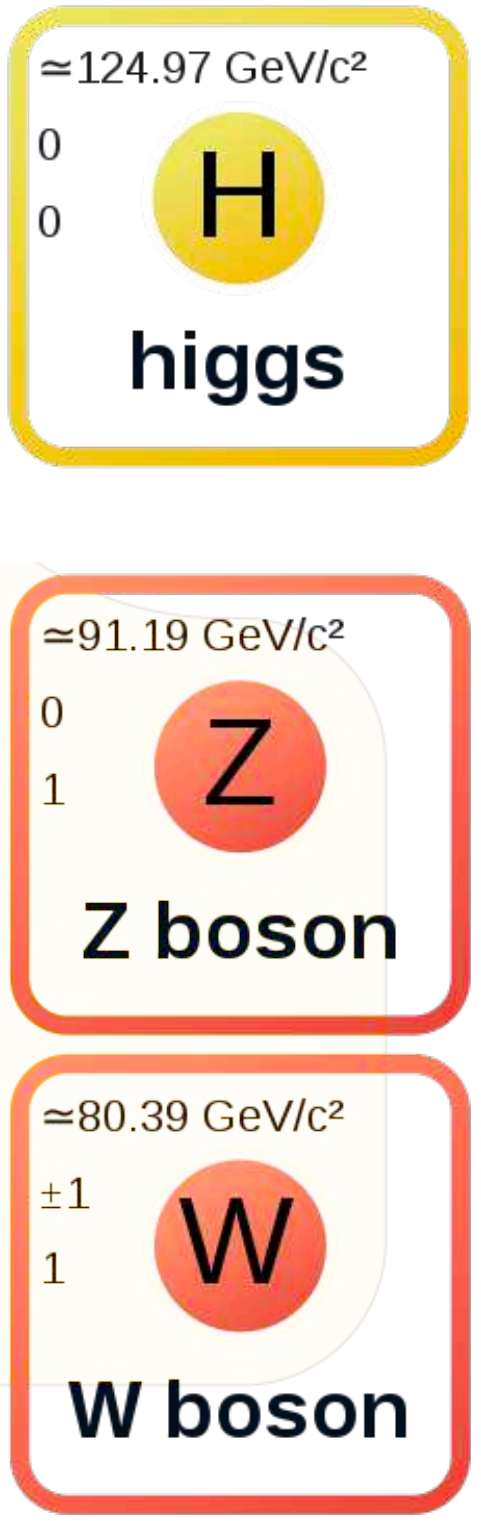


Gluon

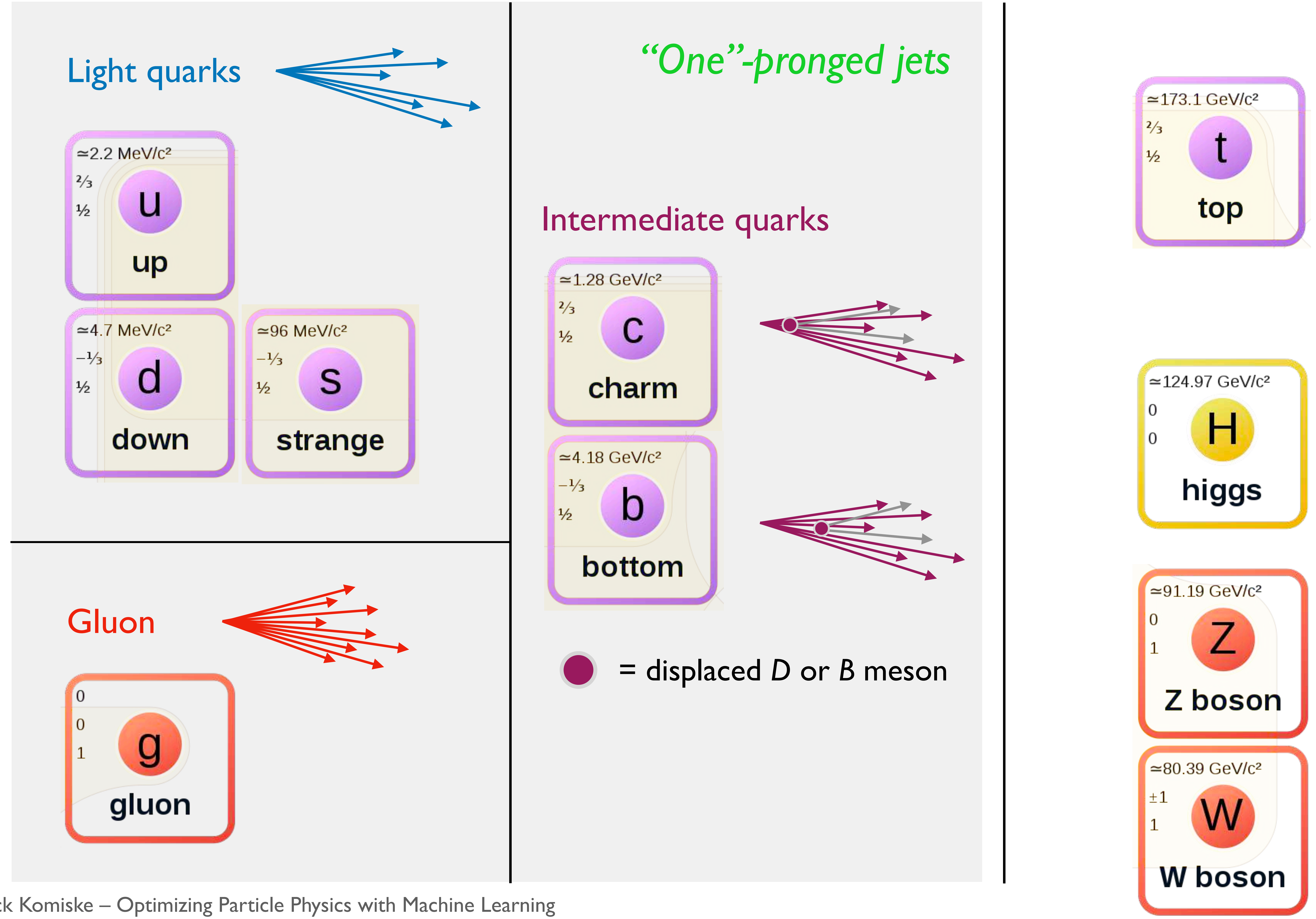
Intermediate quarks



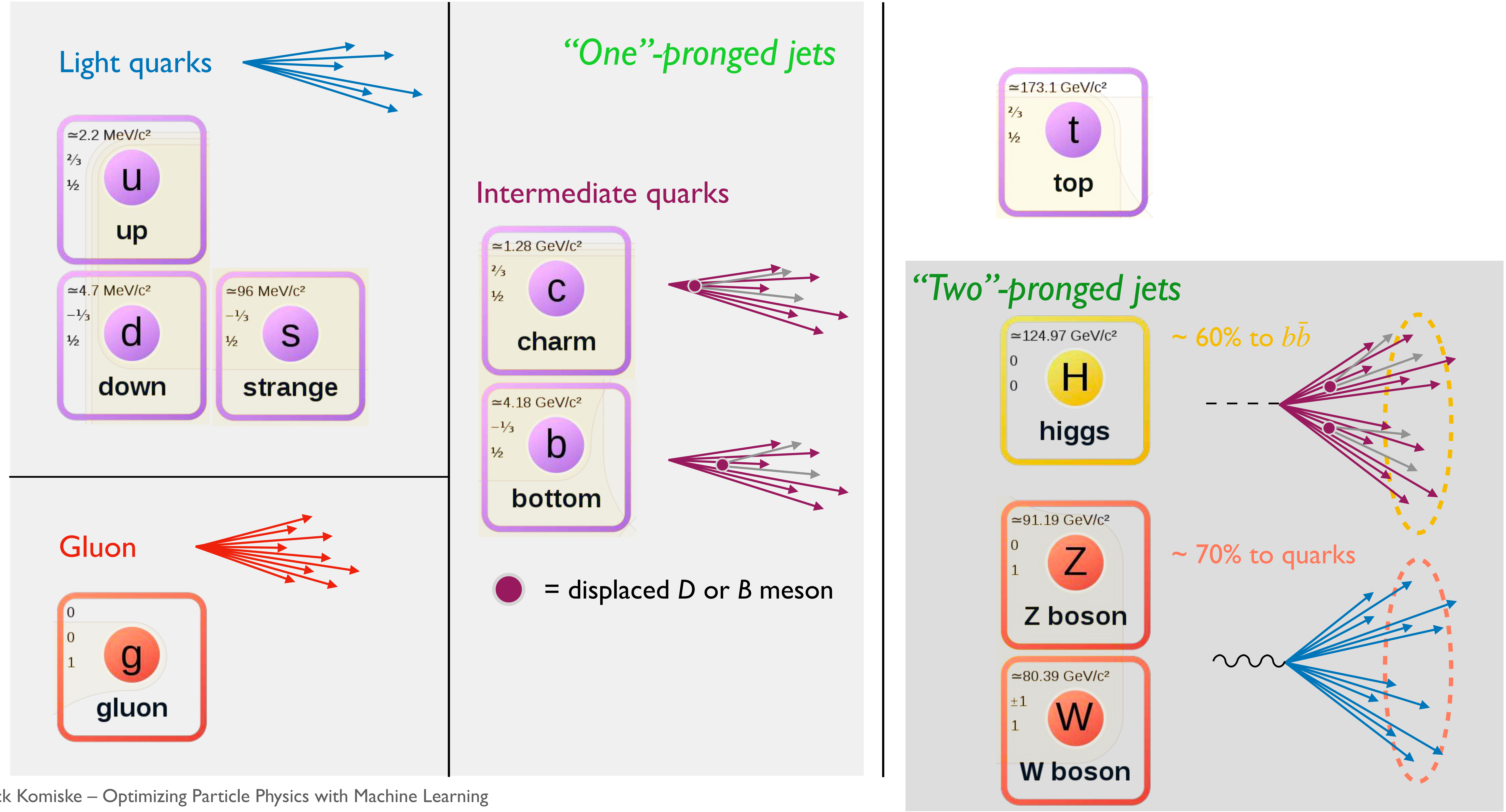




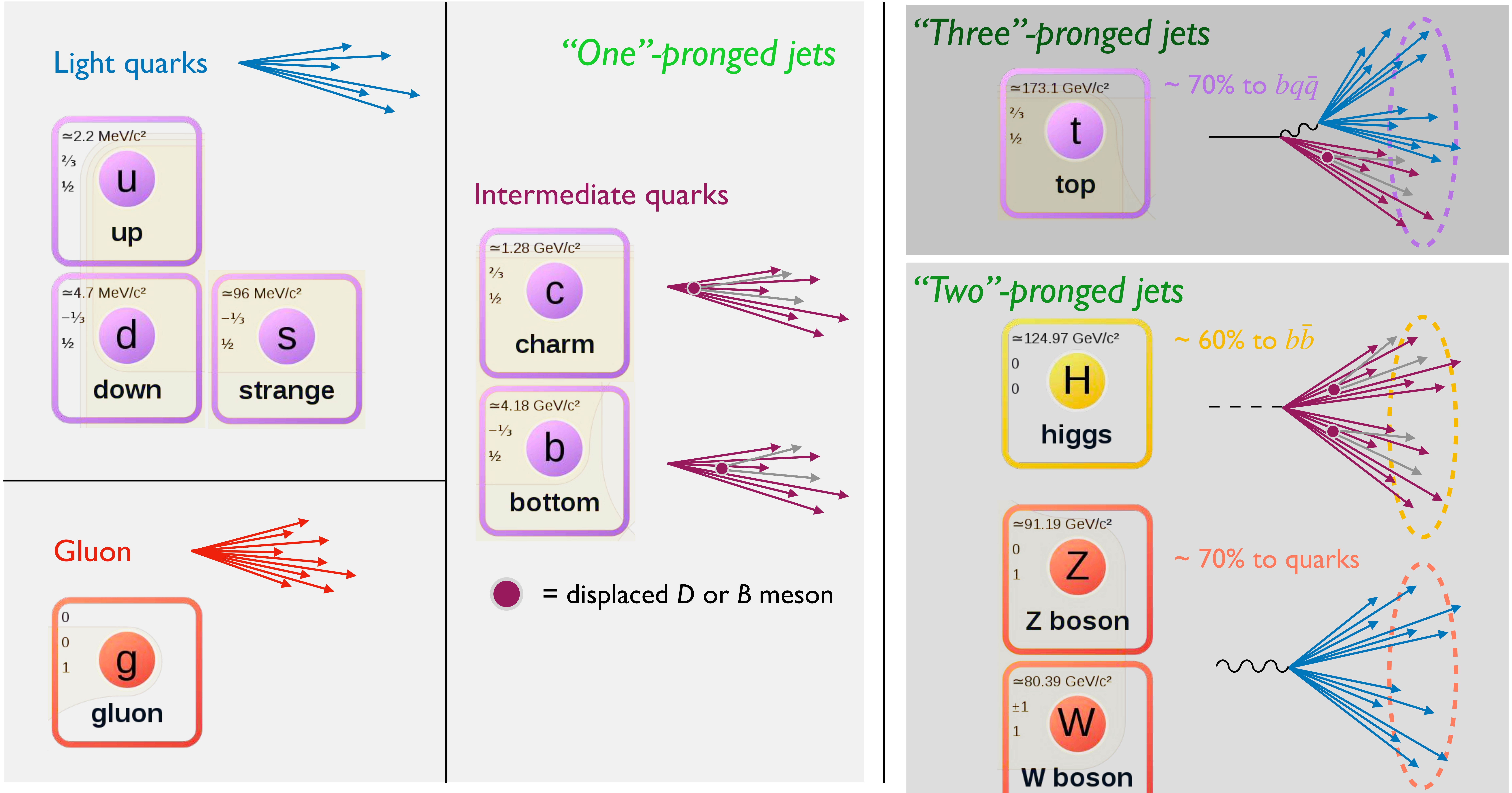
Standard Model of Particle Physics at High-Energies – as Jets



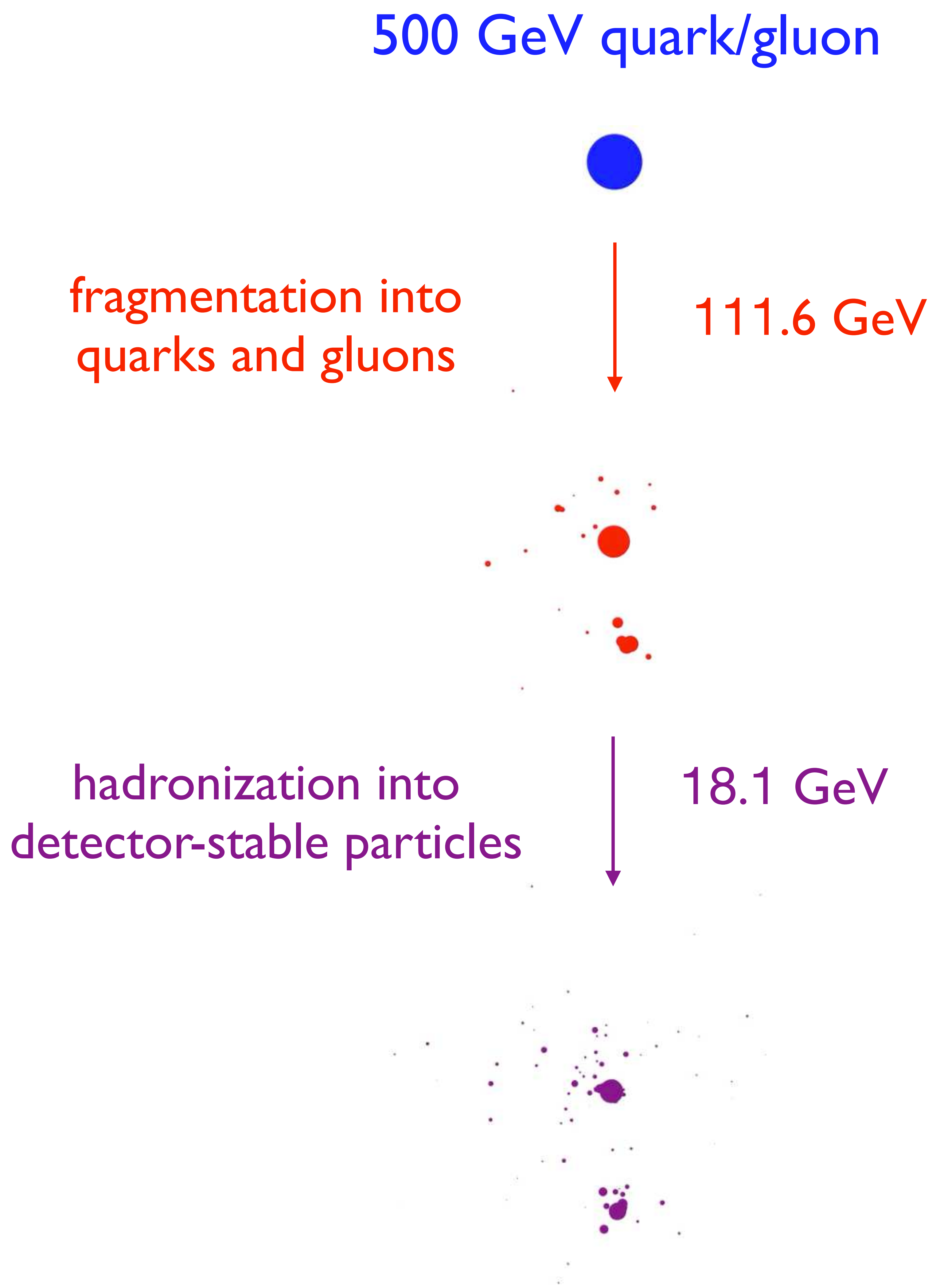
Standard Model of Particle Physics at High-Energies – as Jets

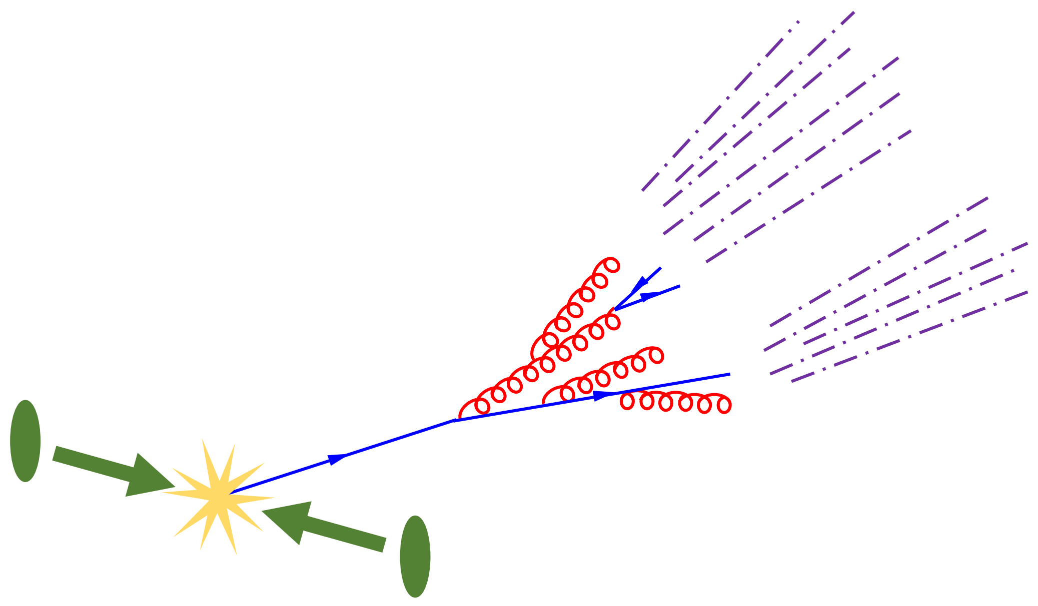


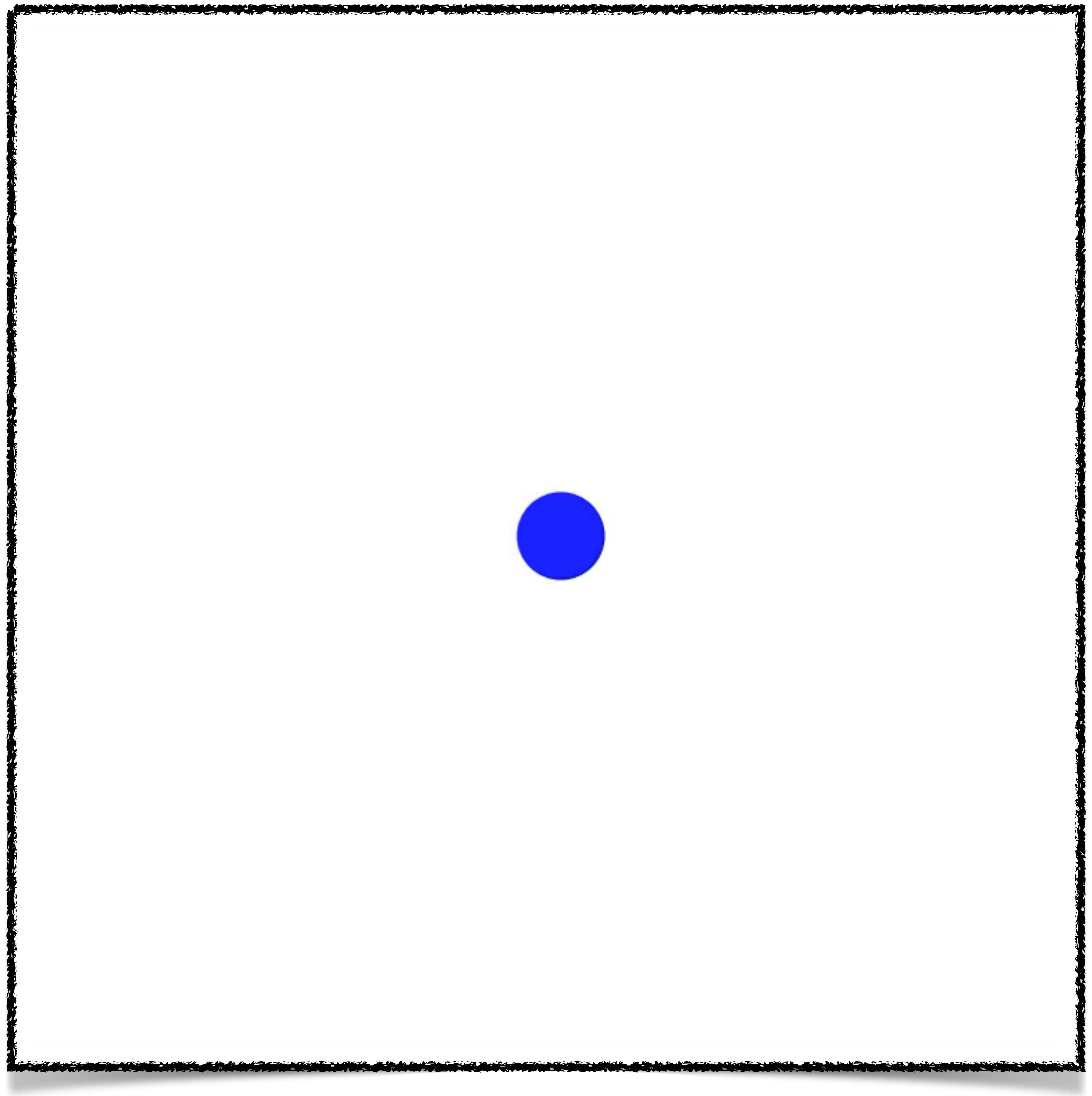
Standard Model of Particle Physics at High-Energies – as Jets



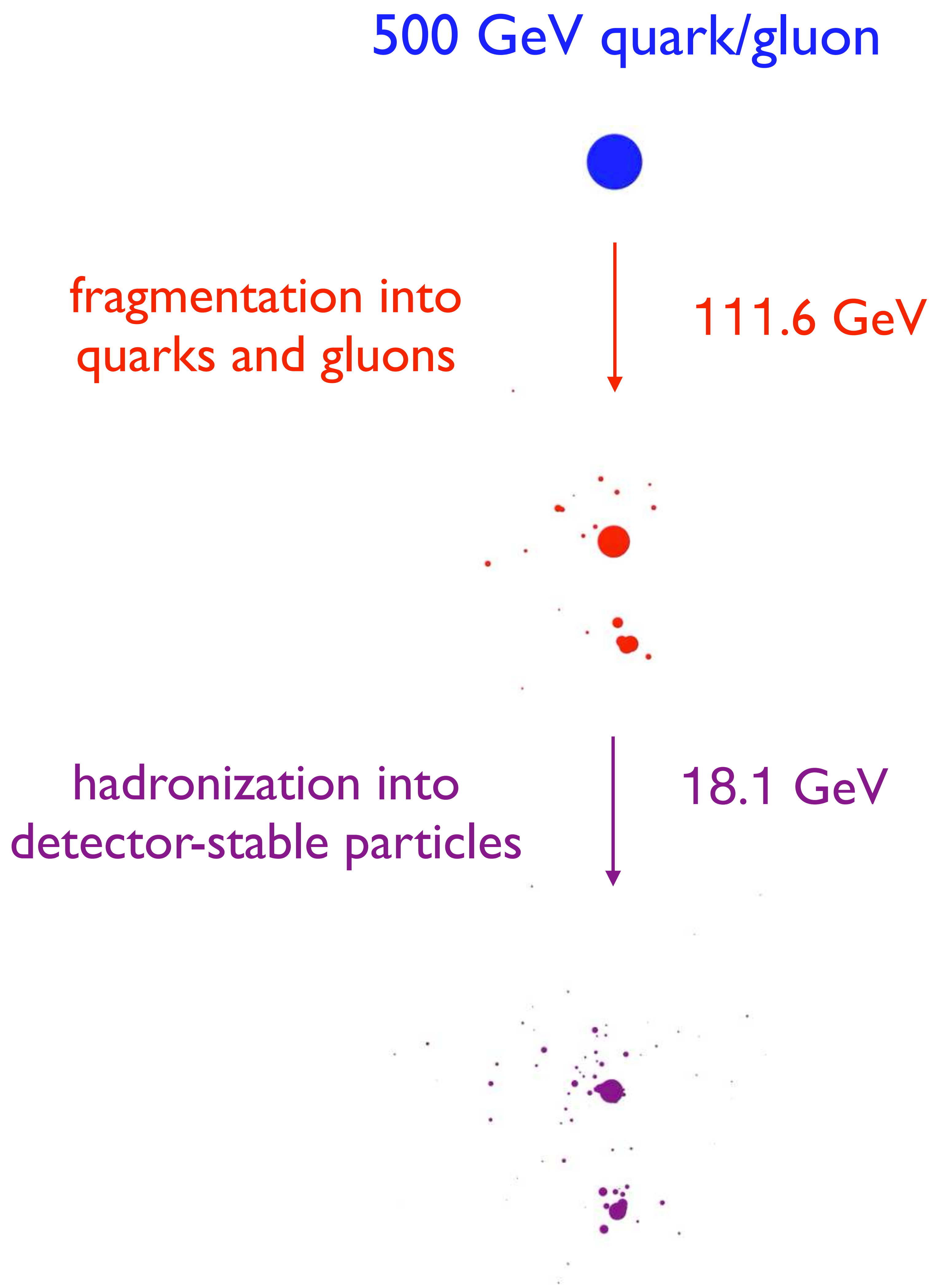
Visualizing Jet Formation – QCD Jets

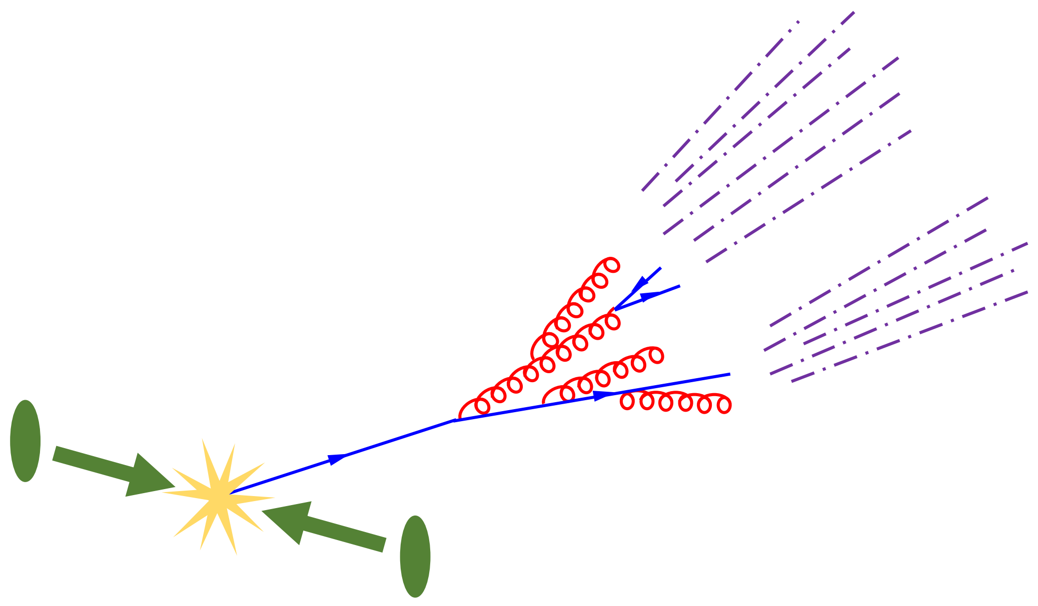


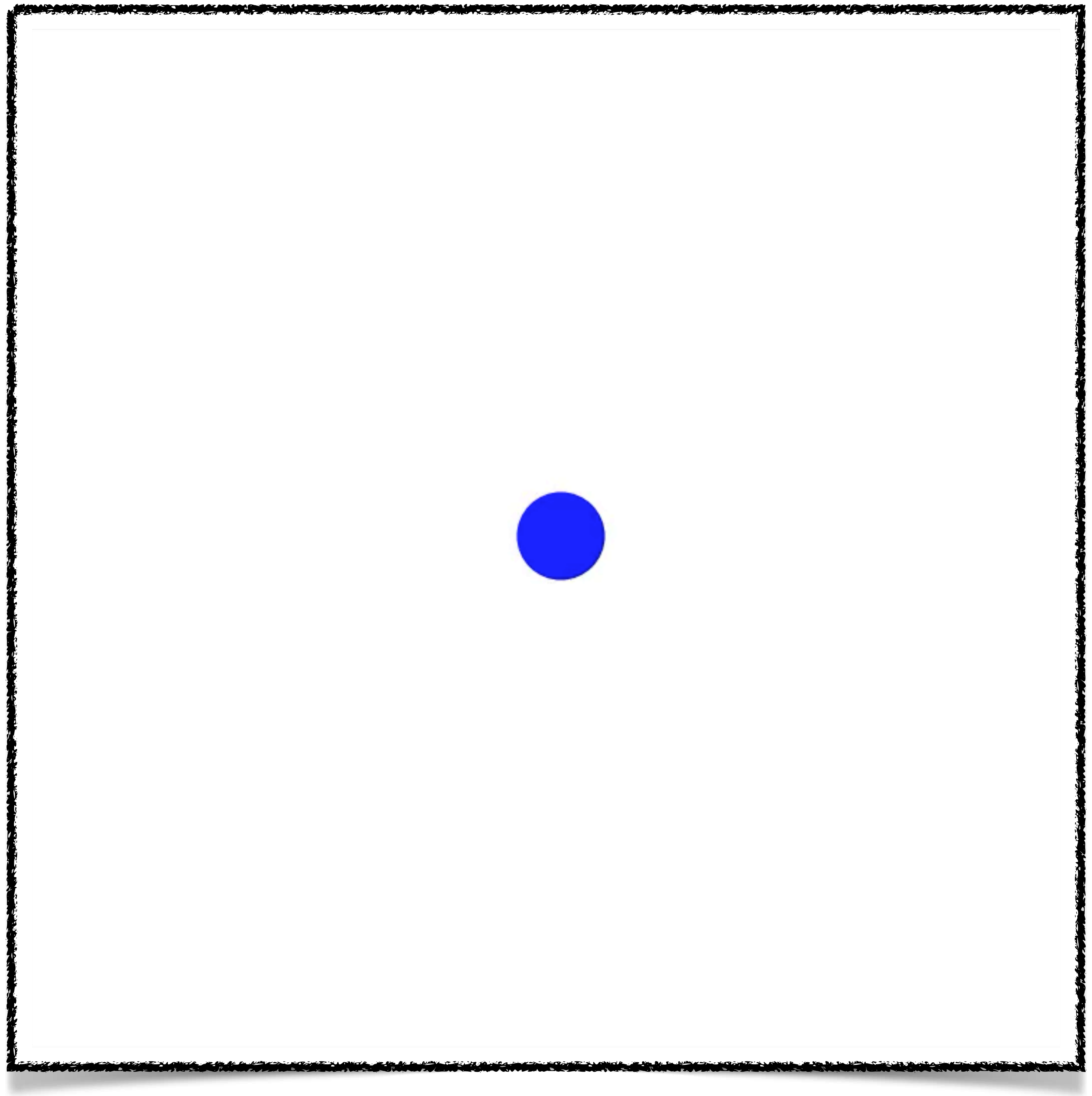




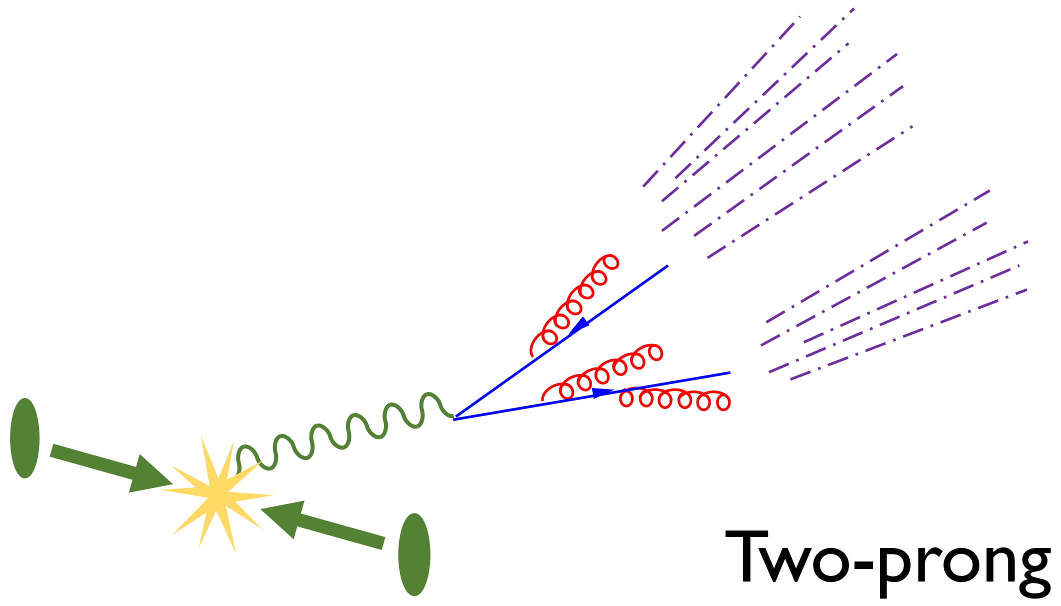
Visualizing Jet Formation – QCD Jets





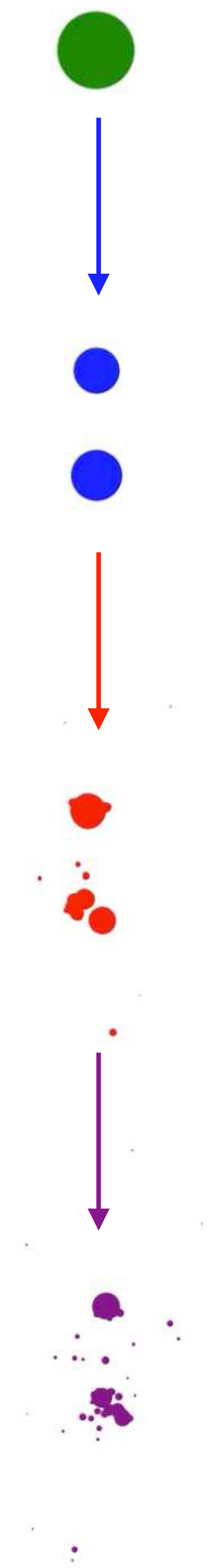


Visualizing Jet Formation – W Jets



Two-prong decay

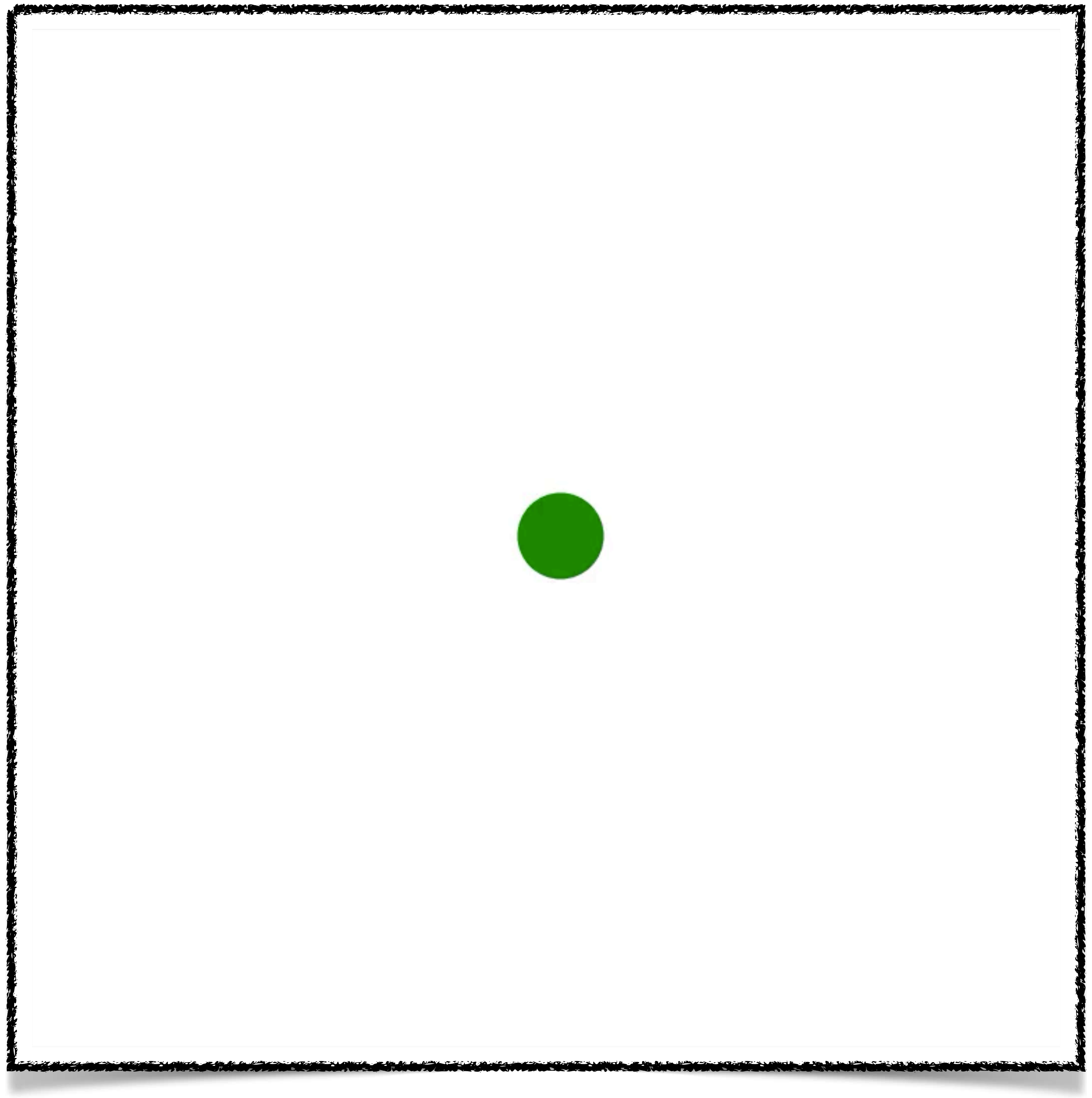
500 GeV W boson



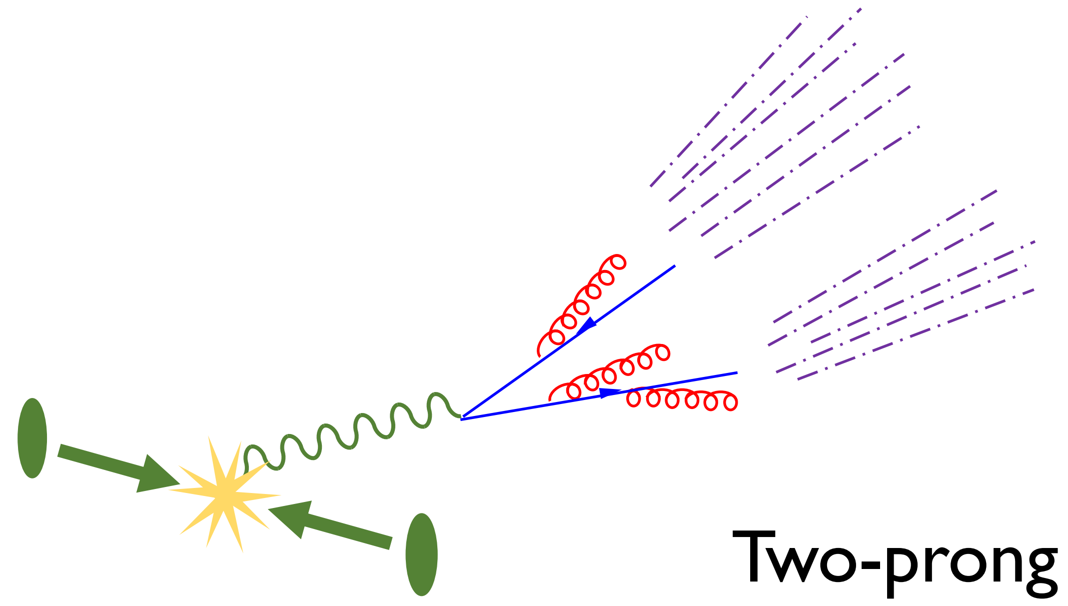
decay to quark/anti-quark — 78.3 GeV

fragmentation into quarks and gluons — 26.3 GeV

hadronization into detector-stable particles — 12.9 GeV



Visualizing Jet Formation – W Jets



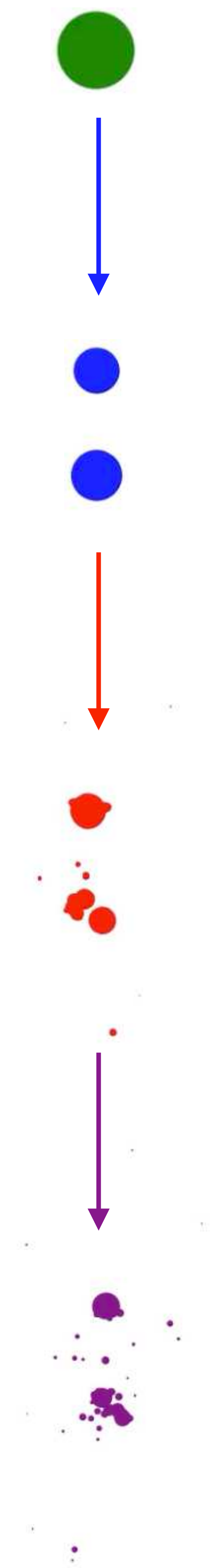
Two-prong decay

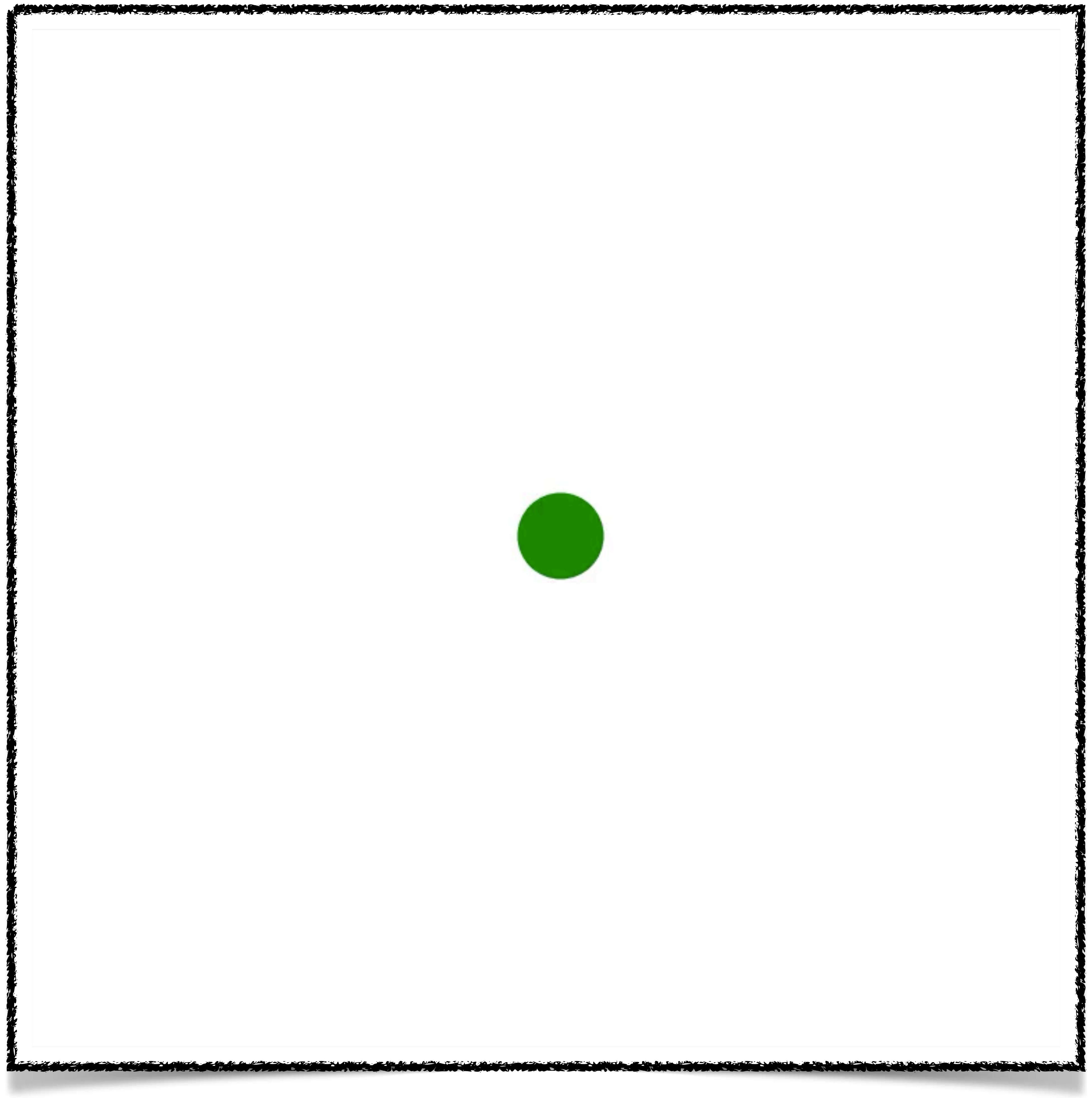
500 GeV *W* boson

decay to quark/anti-quark — 78.3 GeV

fragmentation into quarks and gluons — 26.3 GeV

hadronization into detector-stable particles — 12.9 GeV

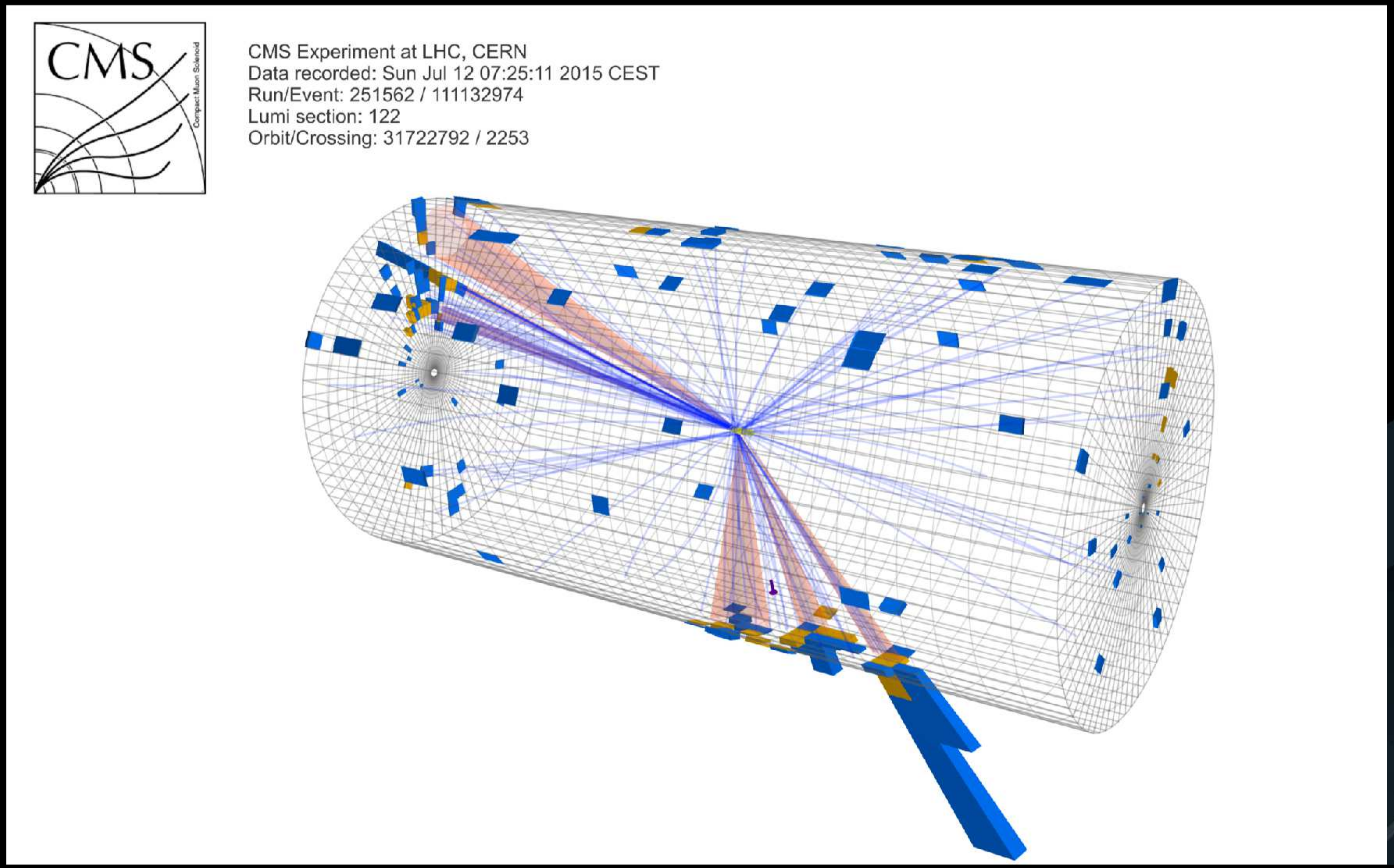




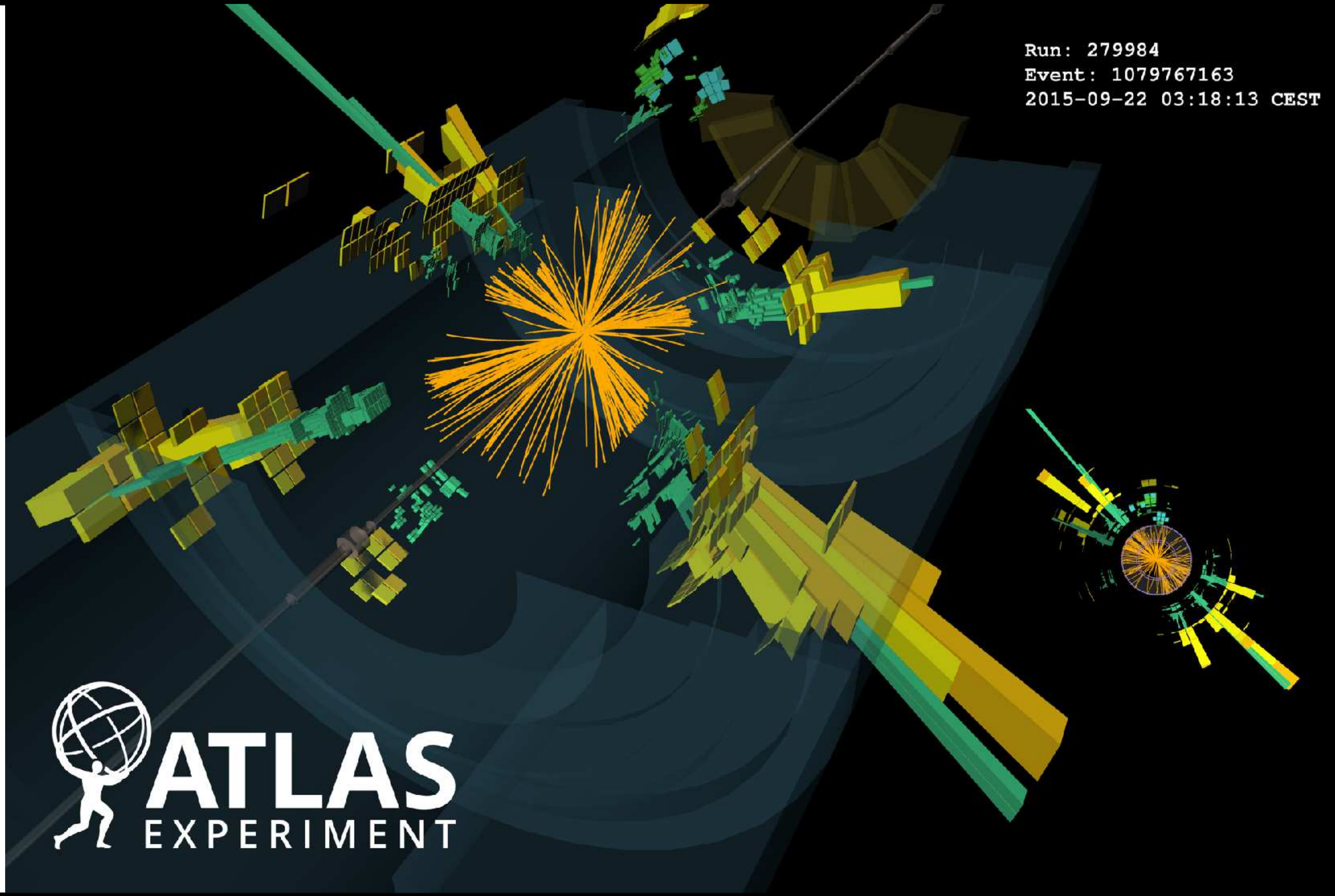
Jet Formation at the LHC

Jets are collimated sprays of particles arising from production of high-energy quarks and gluons

CMS $t\bar{t}$ decaying to 6 light quarks, candidate event



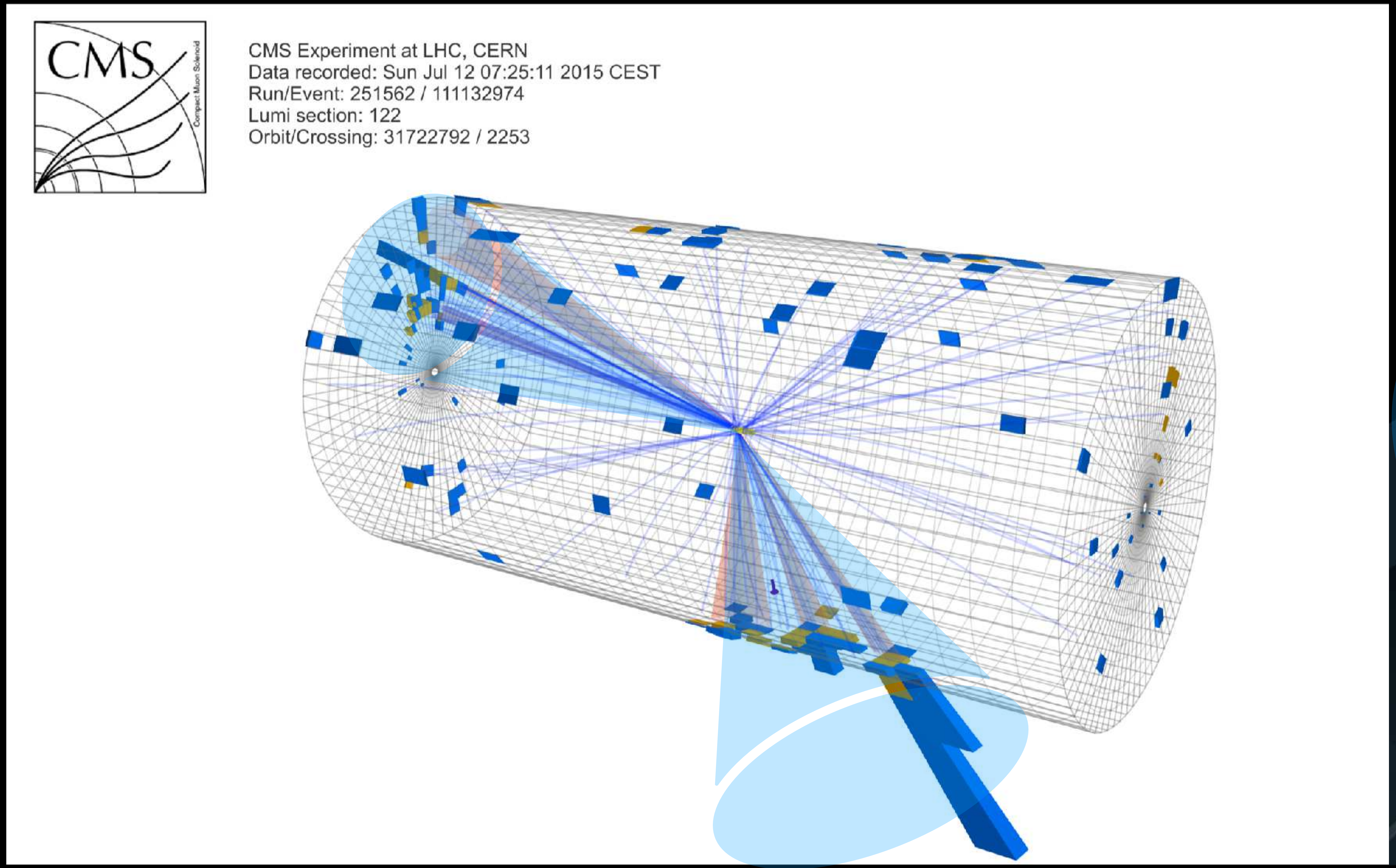
ATLAS multi-jet event



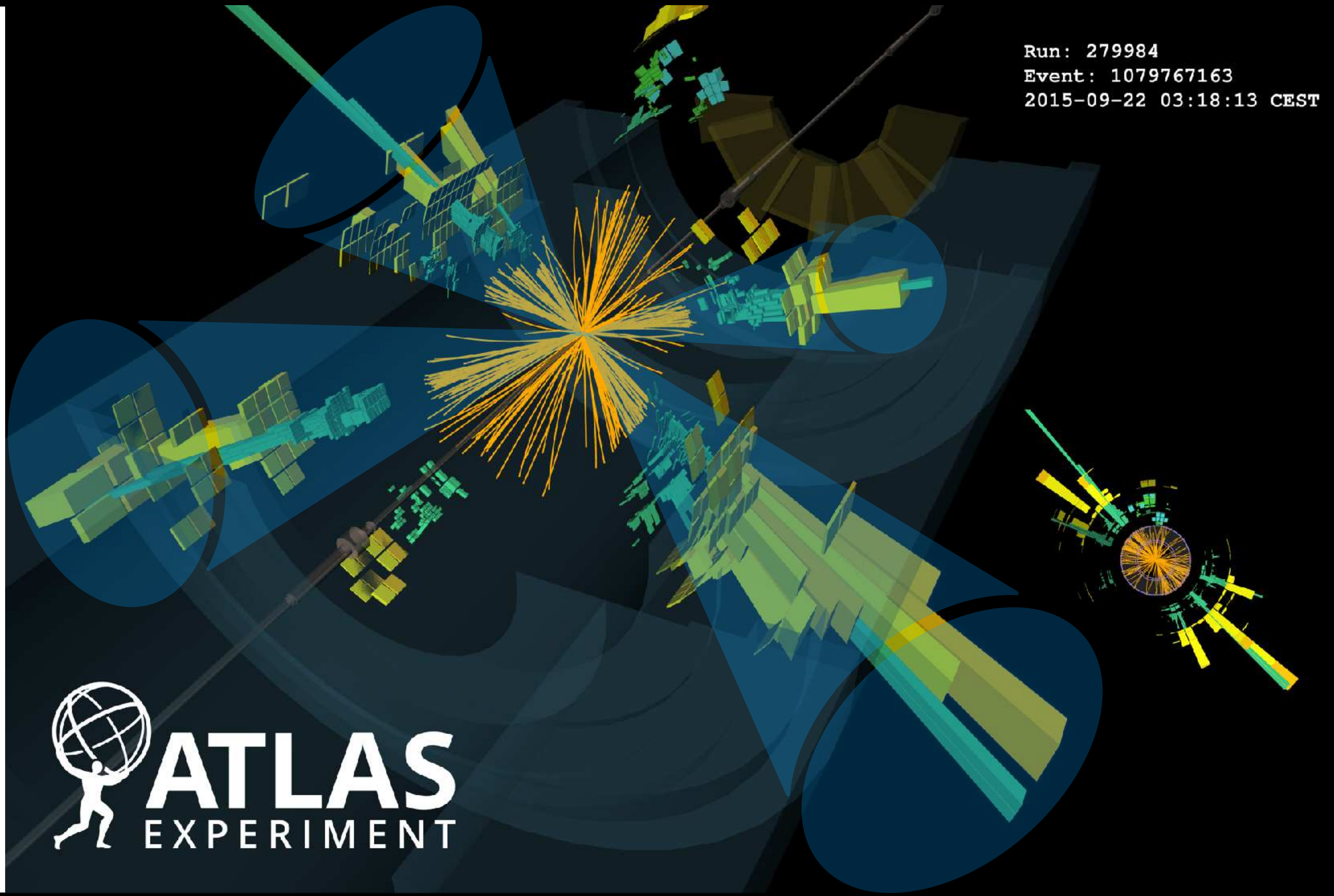
Jet Formation at the LHC

Jets are collimated sprays of particles arising from production of high-energy quarks and gluons

CMS $t\bar{t}$ decaying to 6 light quarks, candidate event



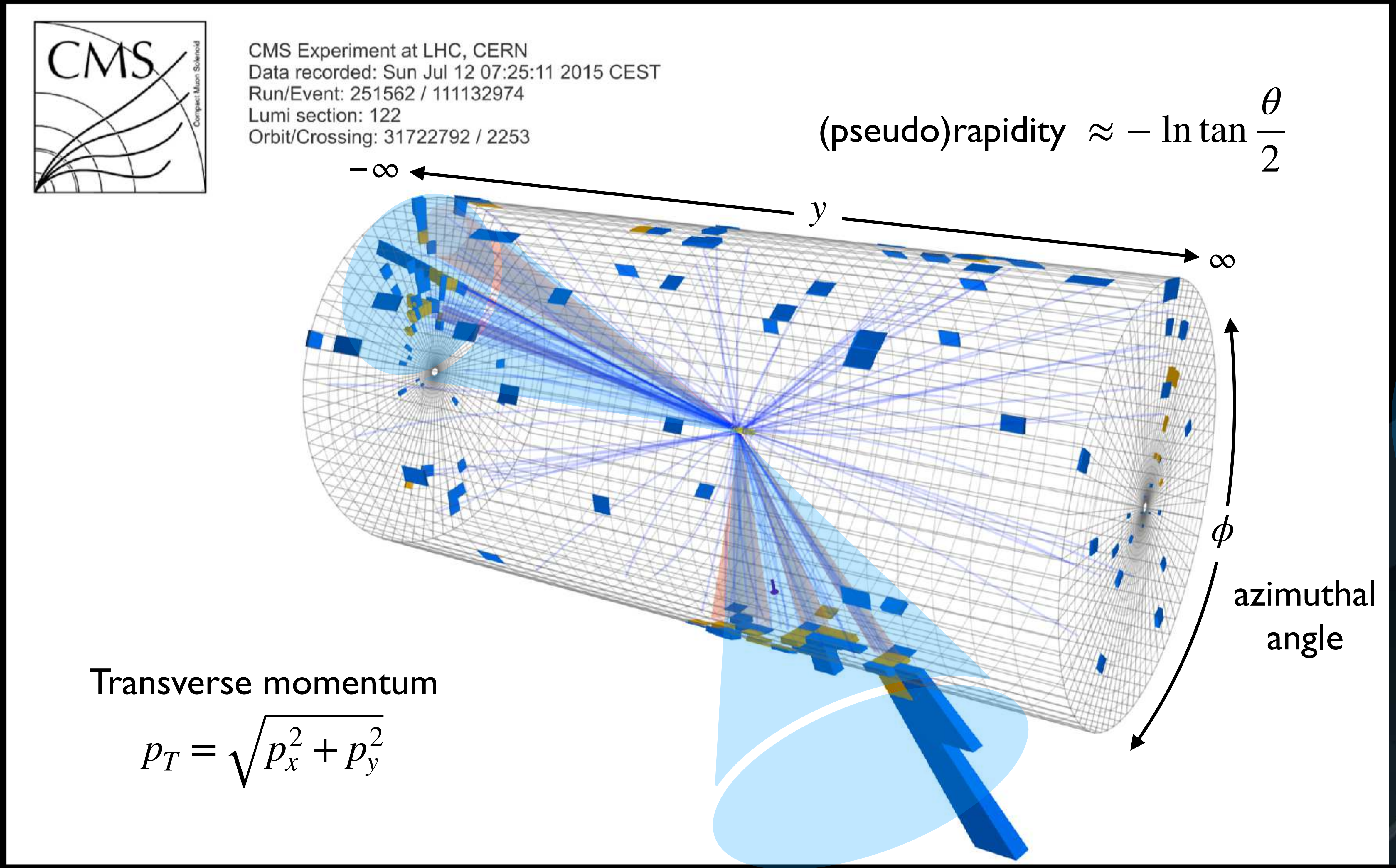
ATLAS multi-jet event



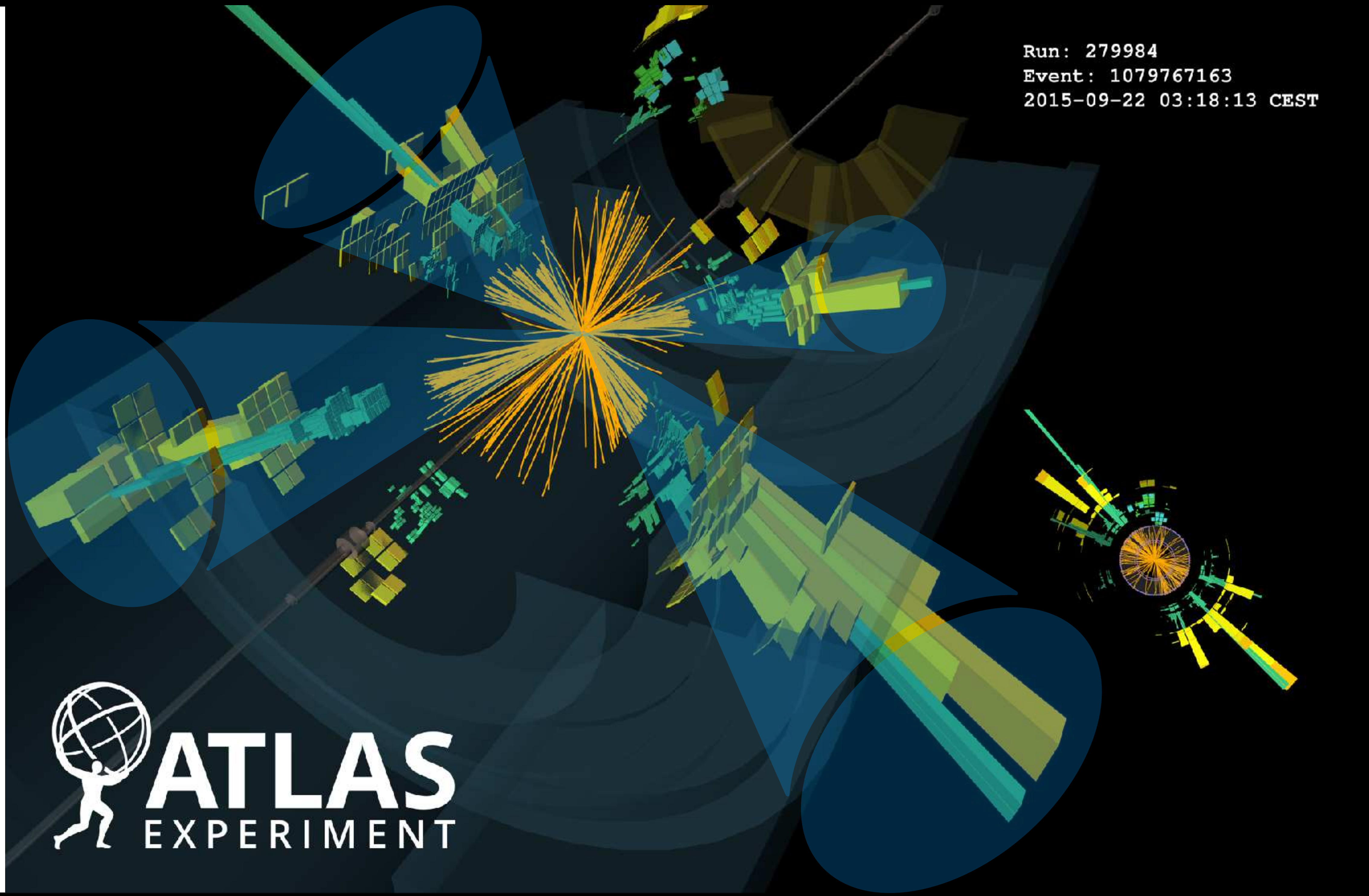
Jet Formation at the LHC

Jets are collimated sprays of particles arising from production of high-energy quarks and gluons

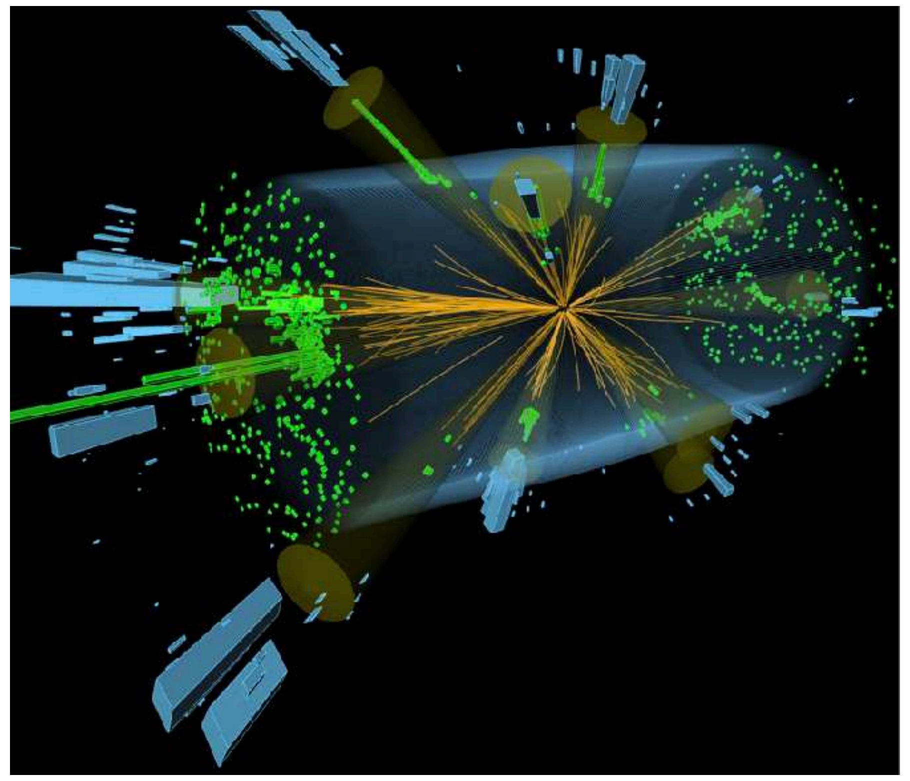
CMS $t\bar{t}$ decaying to 6 light quarks, candidate event



ATLAS multi-jet event



Jets are defined via sequential recombination of particles (hierarchical agglomerative clustering)

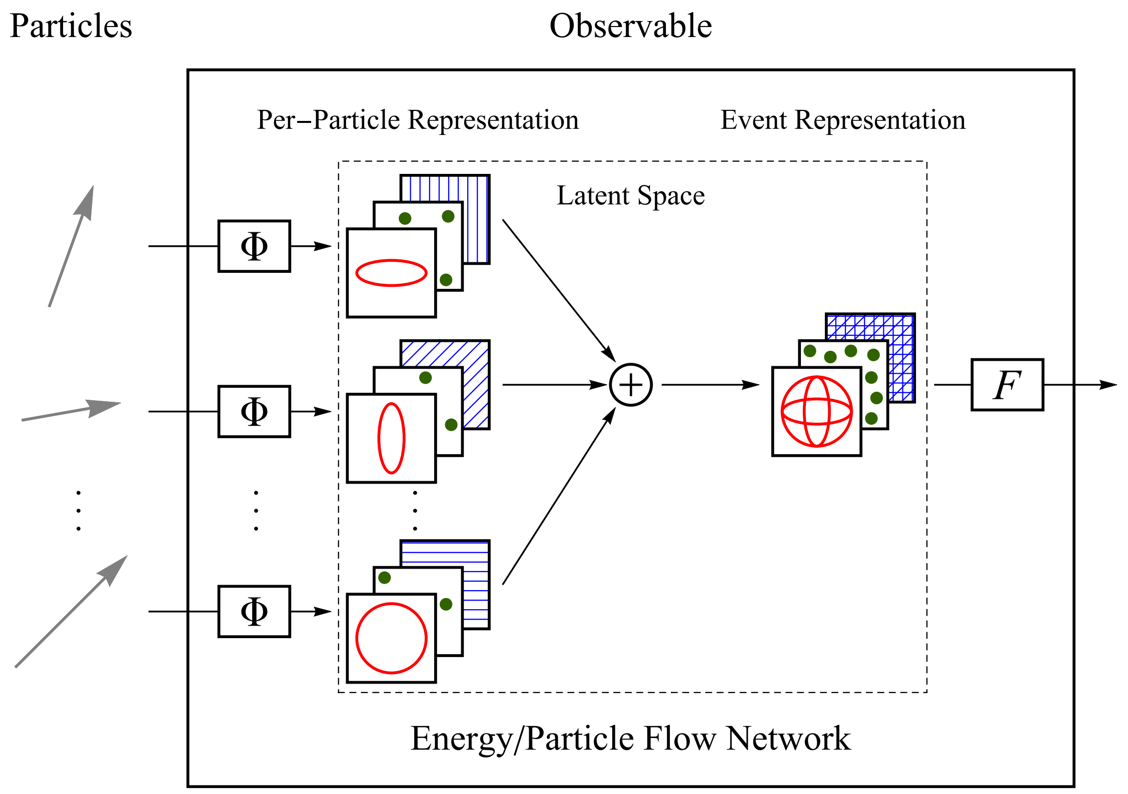


Particle Physics Fundamentals – Jets

Jets are critical to the success of the modern collider program

Architectures for Colliders

Statistical Deconvolution



Particle Physics Fundamentals – Jets

Jets are critical to the success of the modern collider program

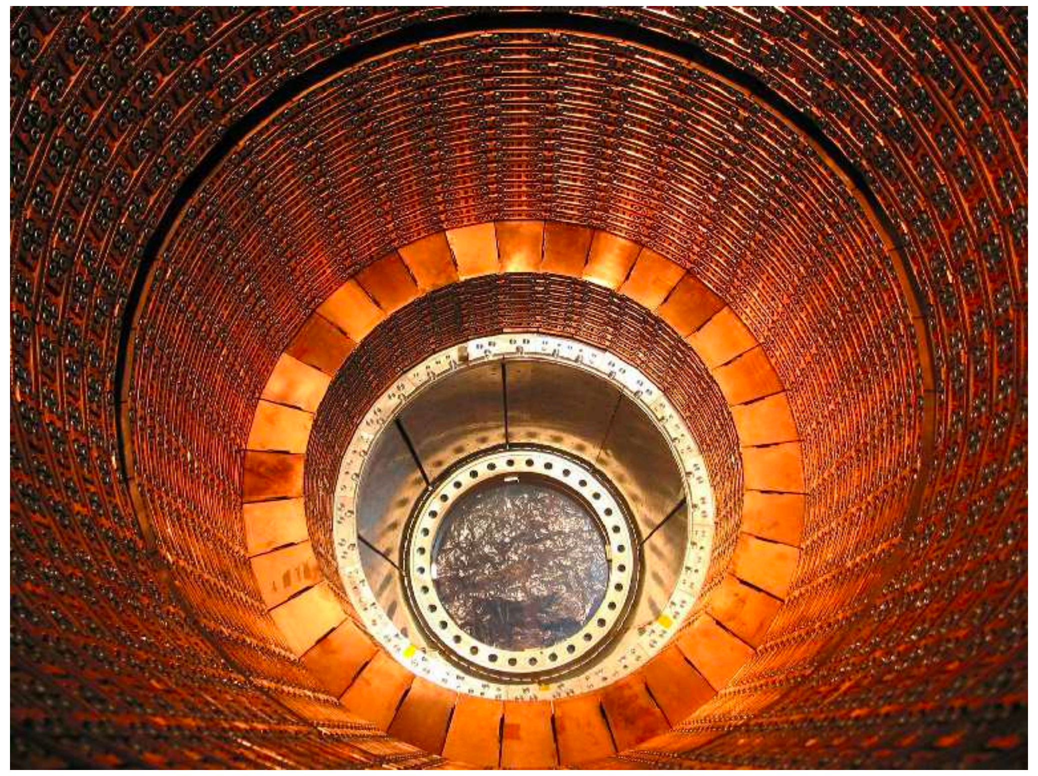
Architectures for Colliders

How do we build a neural network architecture for collider events?

Statistical Deconvolution

Representing Jets as Images

[Cogan, Kagan, Strauss, Schwartzman, JHEP 2015;
de Oliviera, Kagan, Mackey, Nachman, Schwartzman,JHEP 2016;
PTK, Metodiev, Schwartz, JHEP 2017]



as

Take advantage of existing tools for processing images

Pixel intensities ~ transverse momenta of calorimeter cell

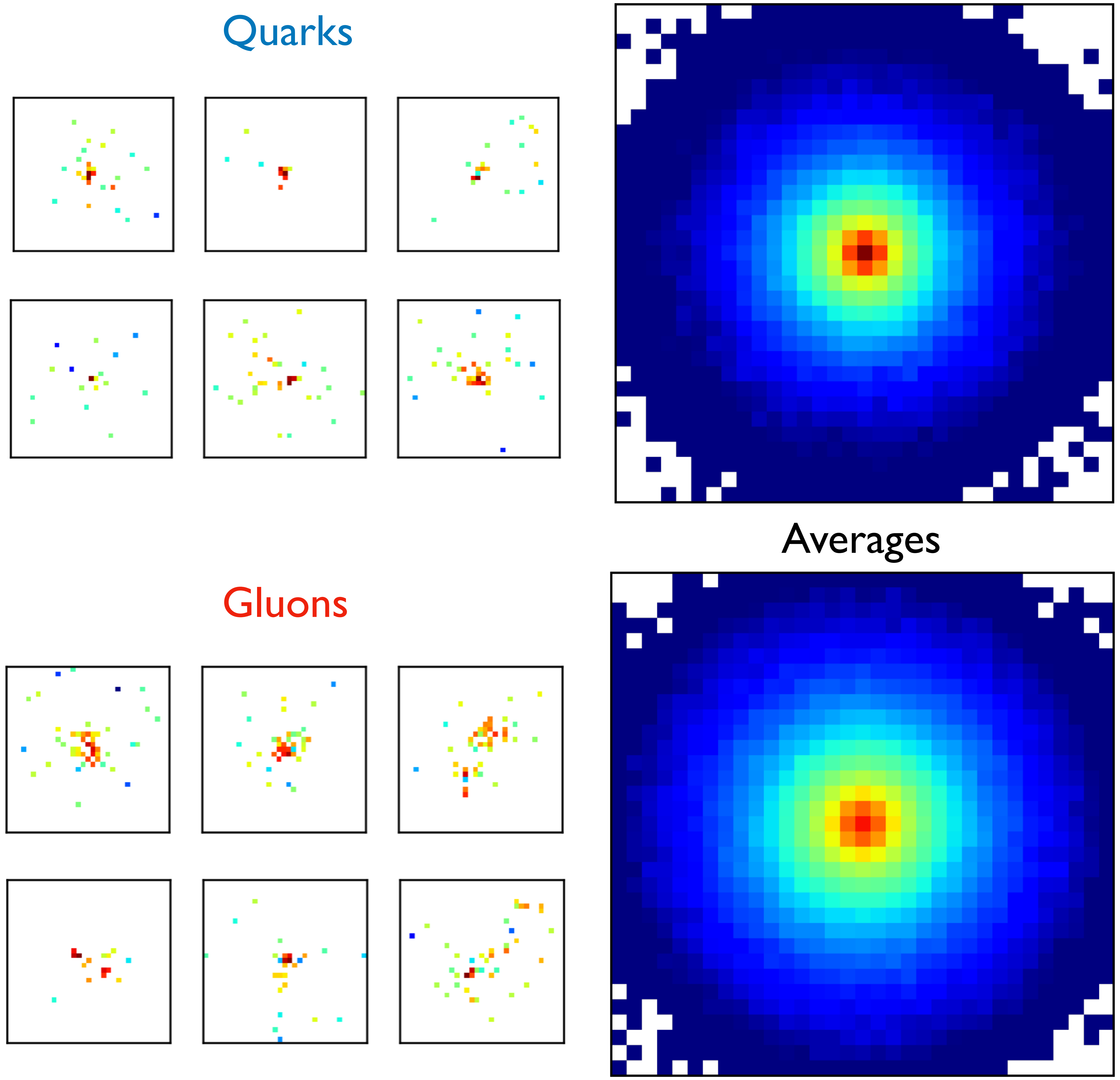
Center on patch of rapidity-azimuth ($y-\phi$) plane containing a jet

"Color" (i.e. multiple channel) images possible, e.g.:

Red: p_T of charged particles

Green: p_T of neutral particles

Blue: charged particle multiplicity

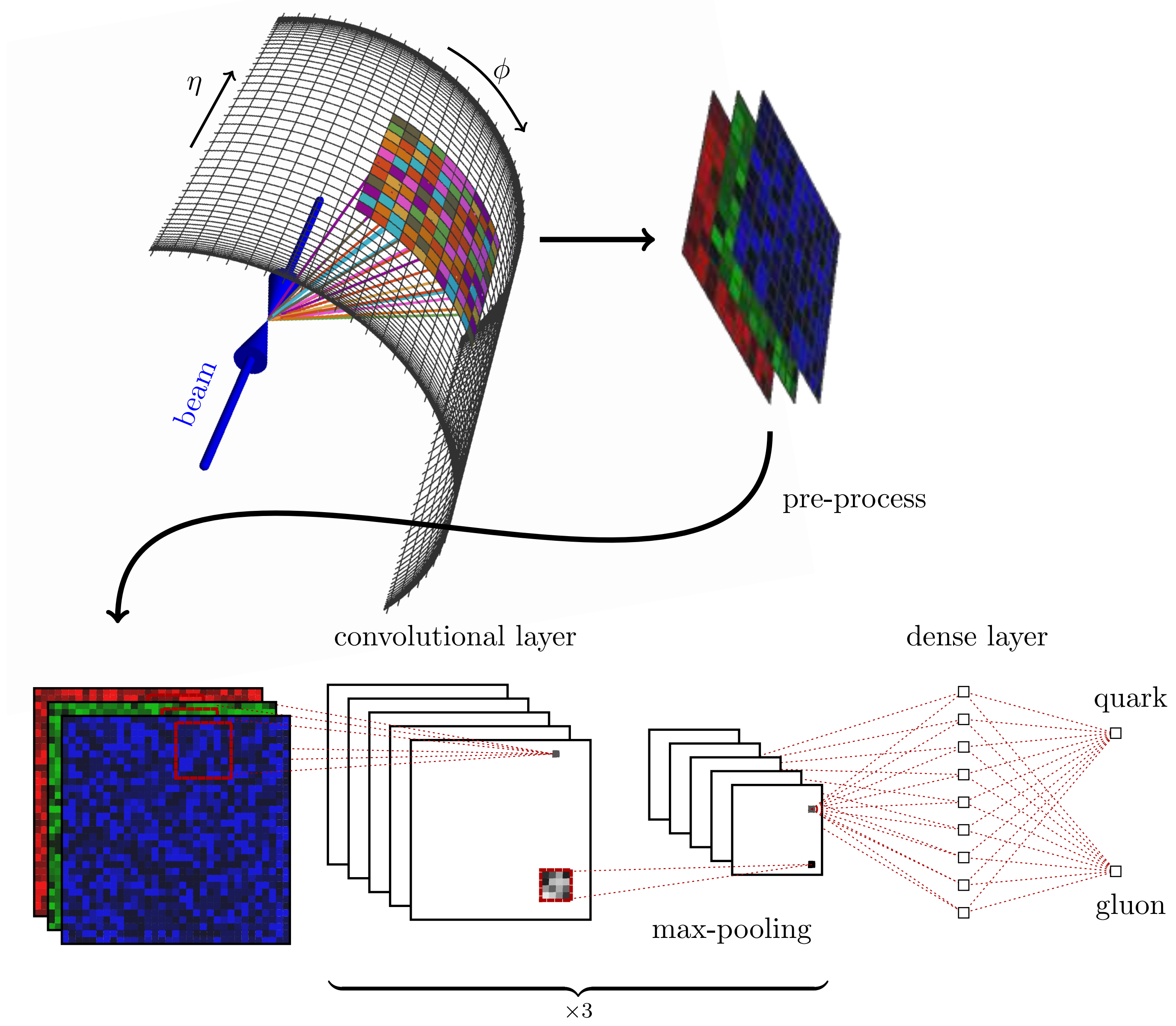


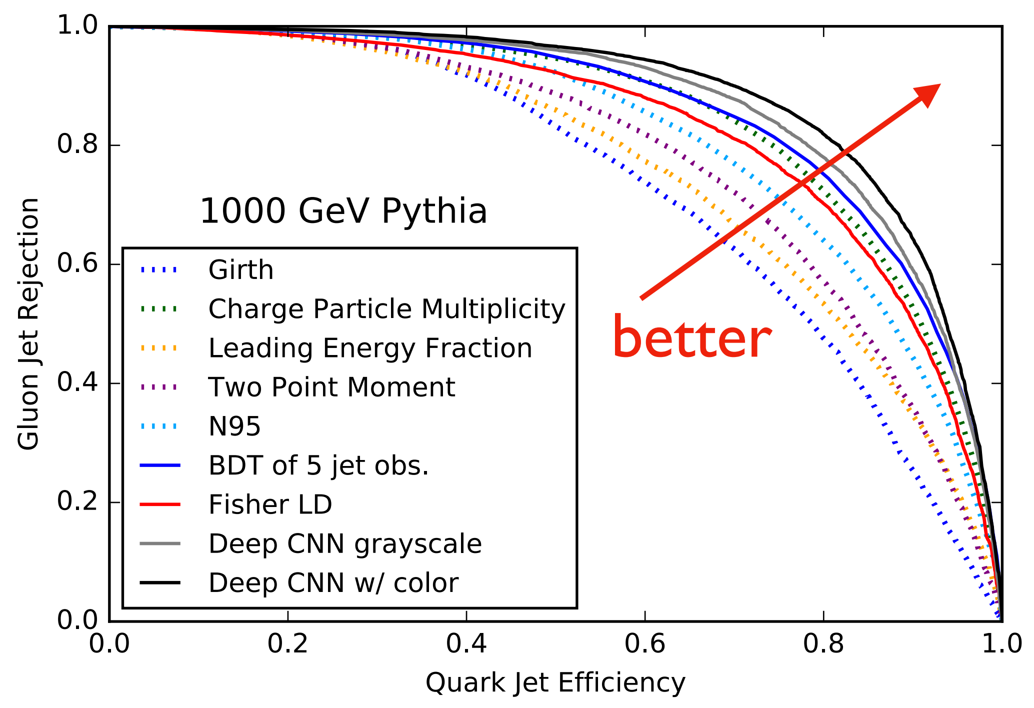
Classification and Regression with Jet Images and Convolutional NNs

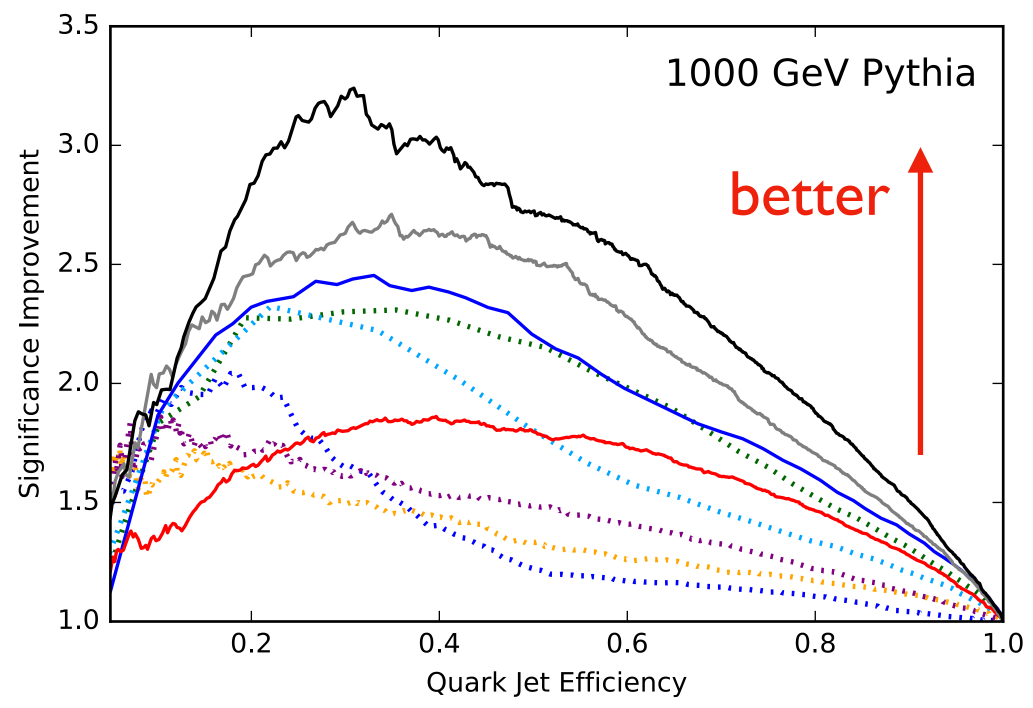
Deep Learning in Color

[PTK, Metodiev, Schwartz, JHEP 2017]

Quark and gluon jet classification with multi-channel jet images



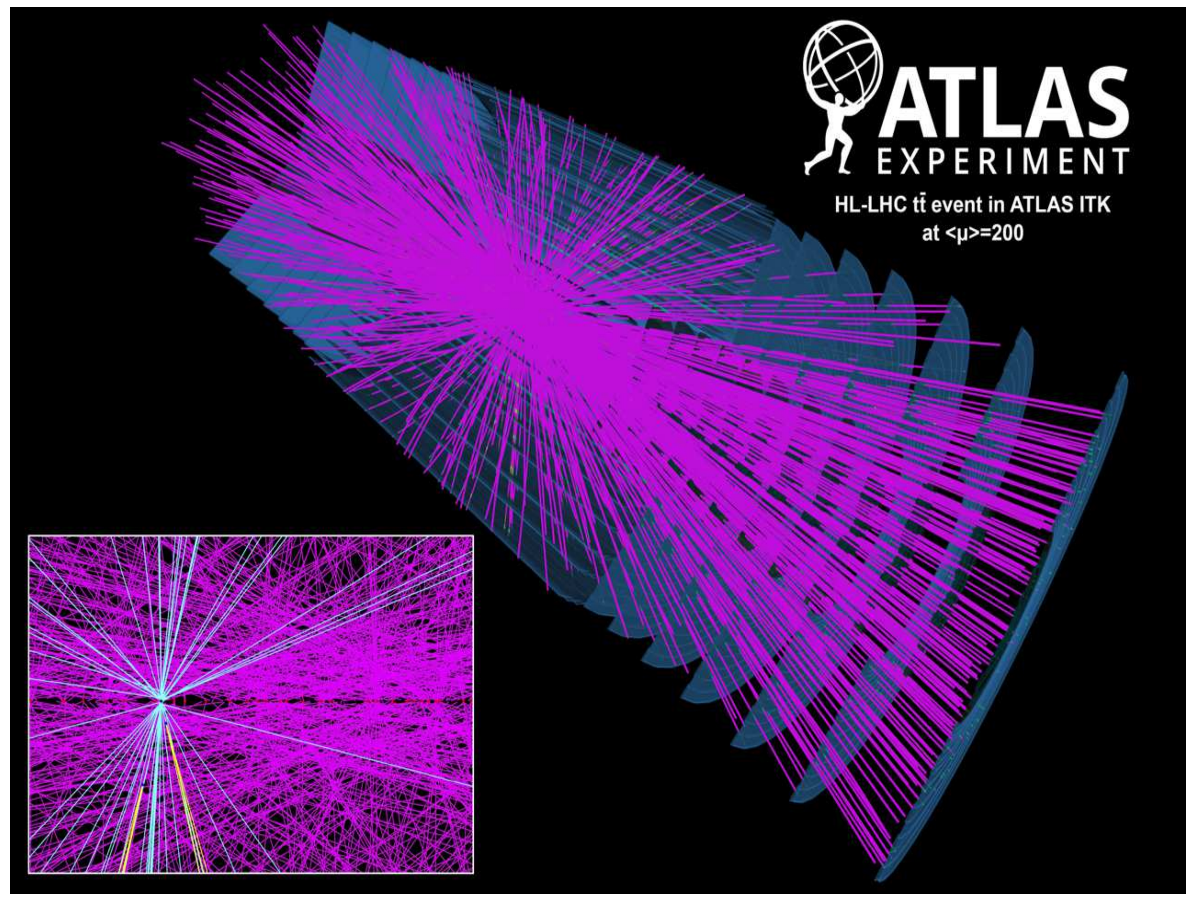


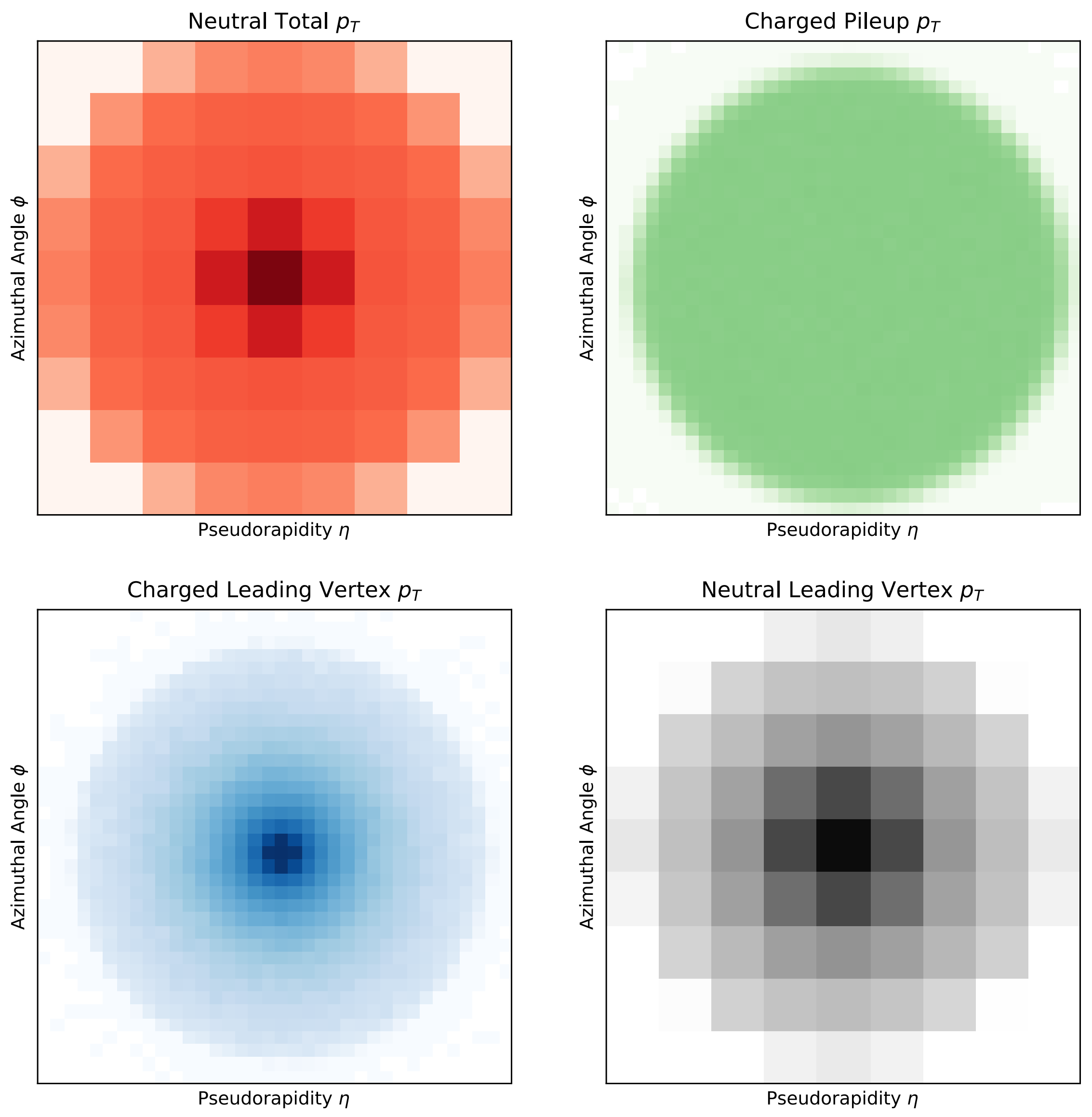


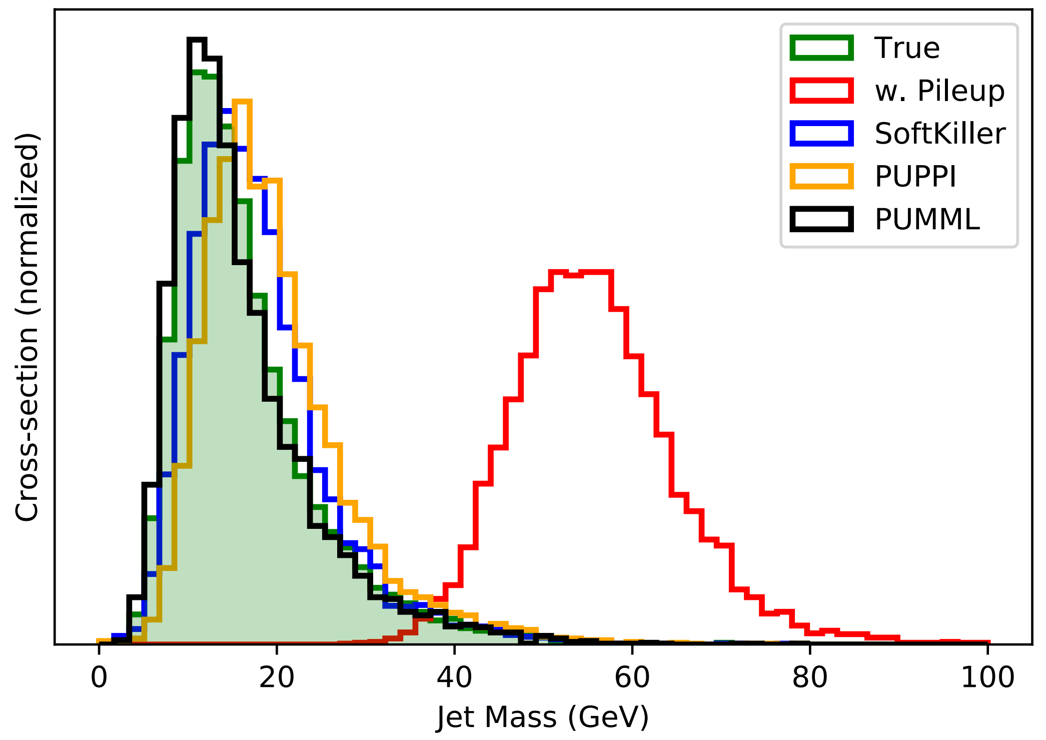
Pileup Mitigation with Machine Learning (PUMML)

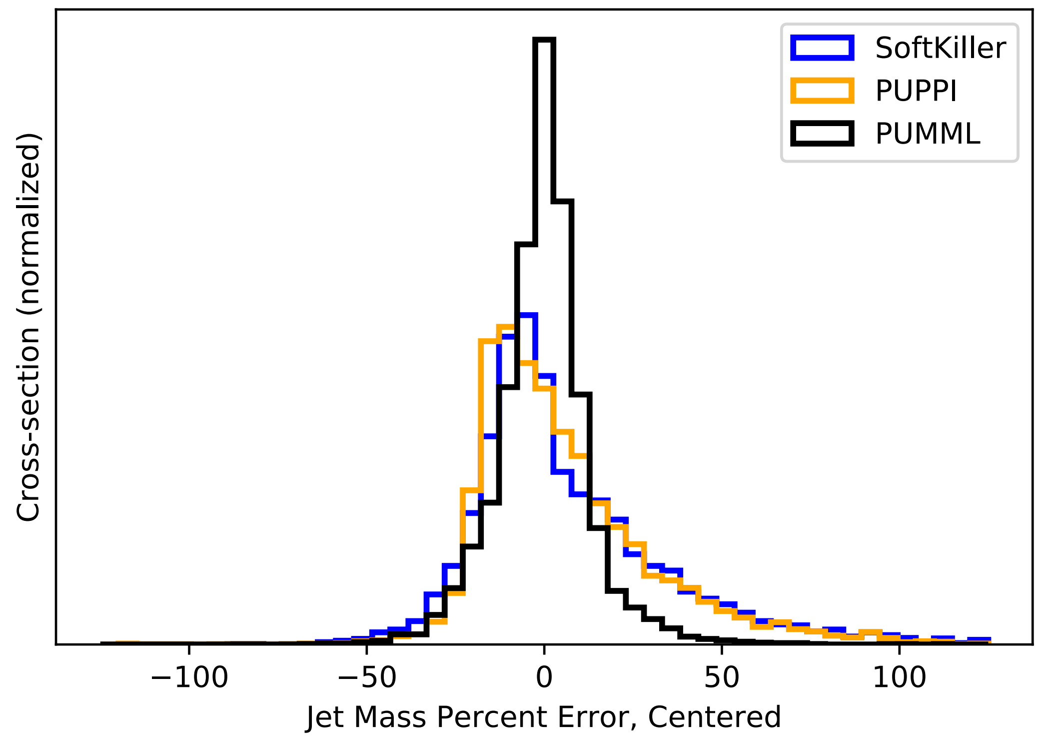
[PTK, Metodiev, Nachman, Schwartz, JHEP 2017]

Pileup removal via regression to the "leading vertex" jet image









Better Neural Network Architectures for Particle Physics

Maximally appropriate ML architectures respect symmetries of the underlying data

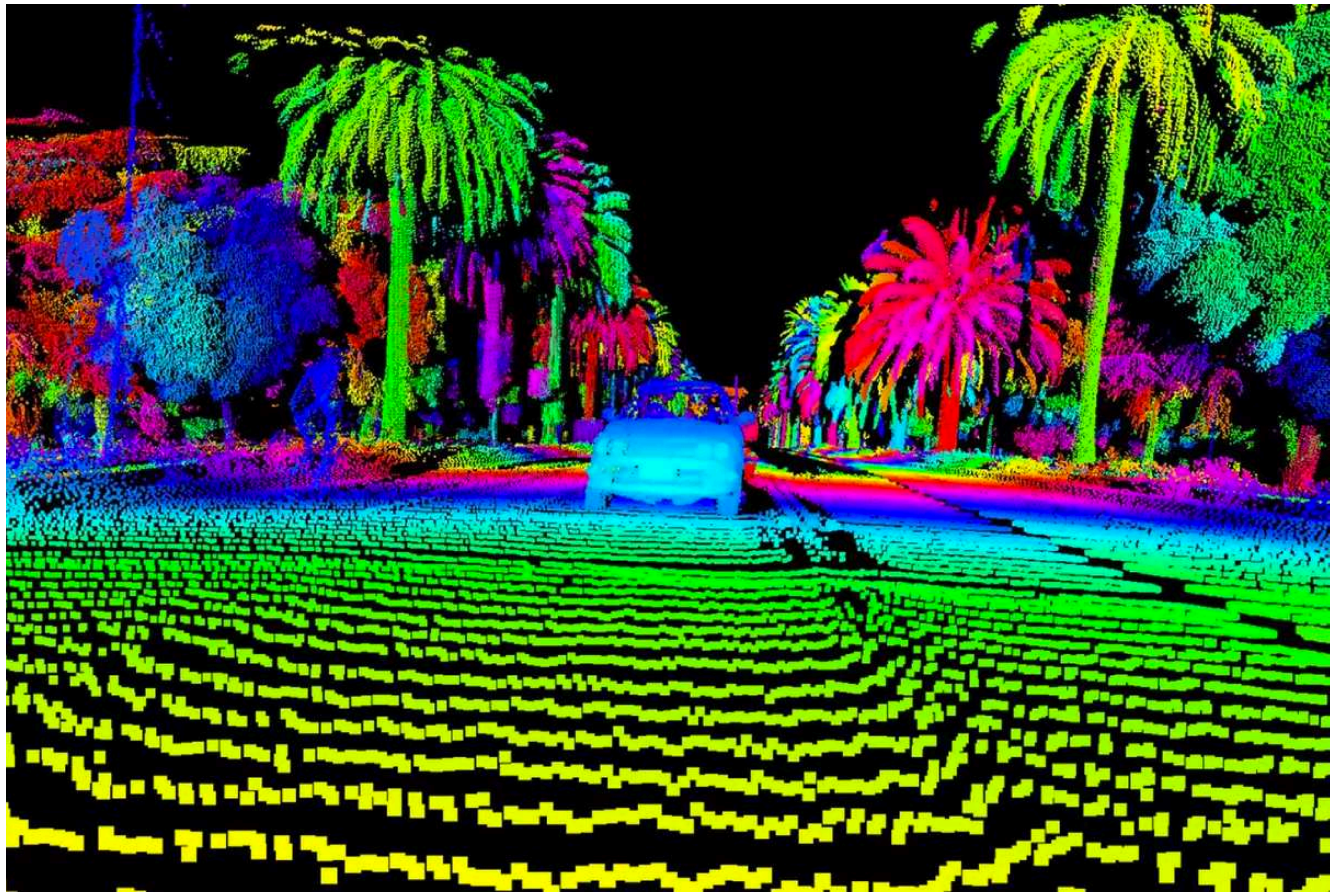
Particle physics events are naturally "point clouds"

Better Neural Network Architectures for Particle Physics

Maximally appropriate ML architectures respect symmetries of the underlying data

Particle physics events are naturally "point clouds"

Point cloud: "A set of data points in space" – Wikipedia



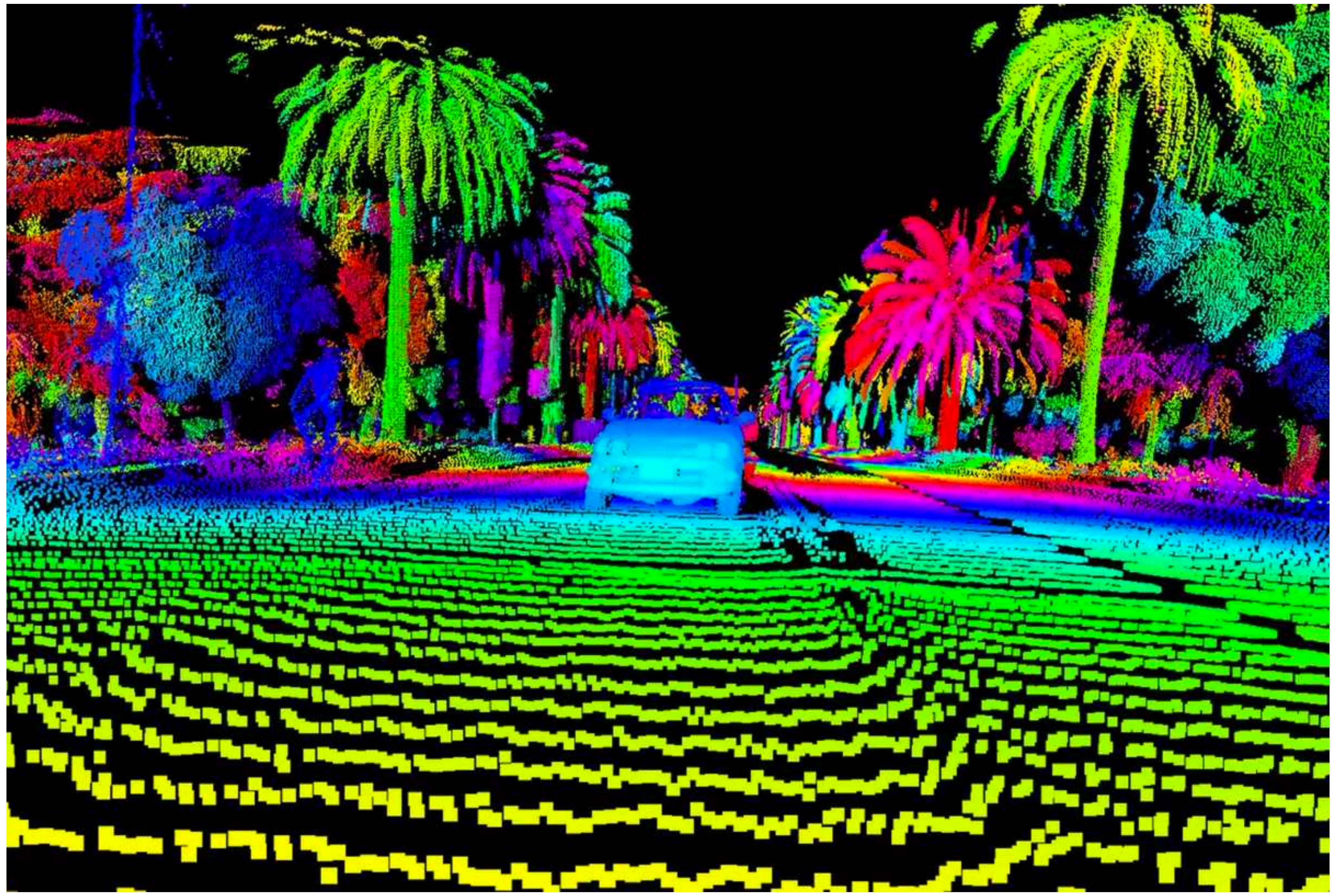
LIDAR data from self-driving car sensor

Better Neural Network Architectures for Particle Physics

Maximally appropriate ML architectures respect symmetries of the underlying data

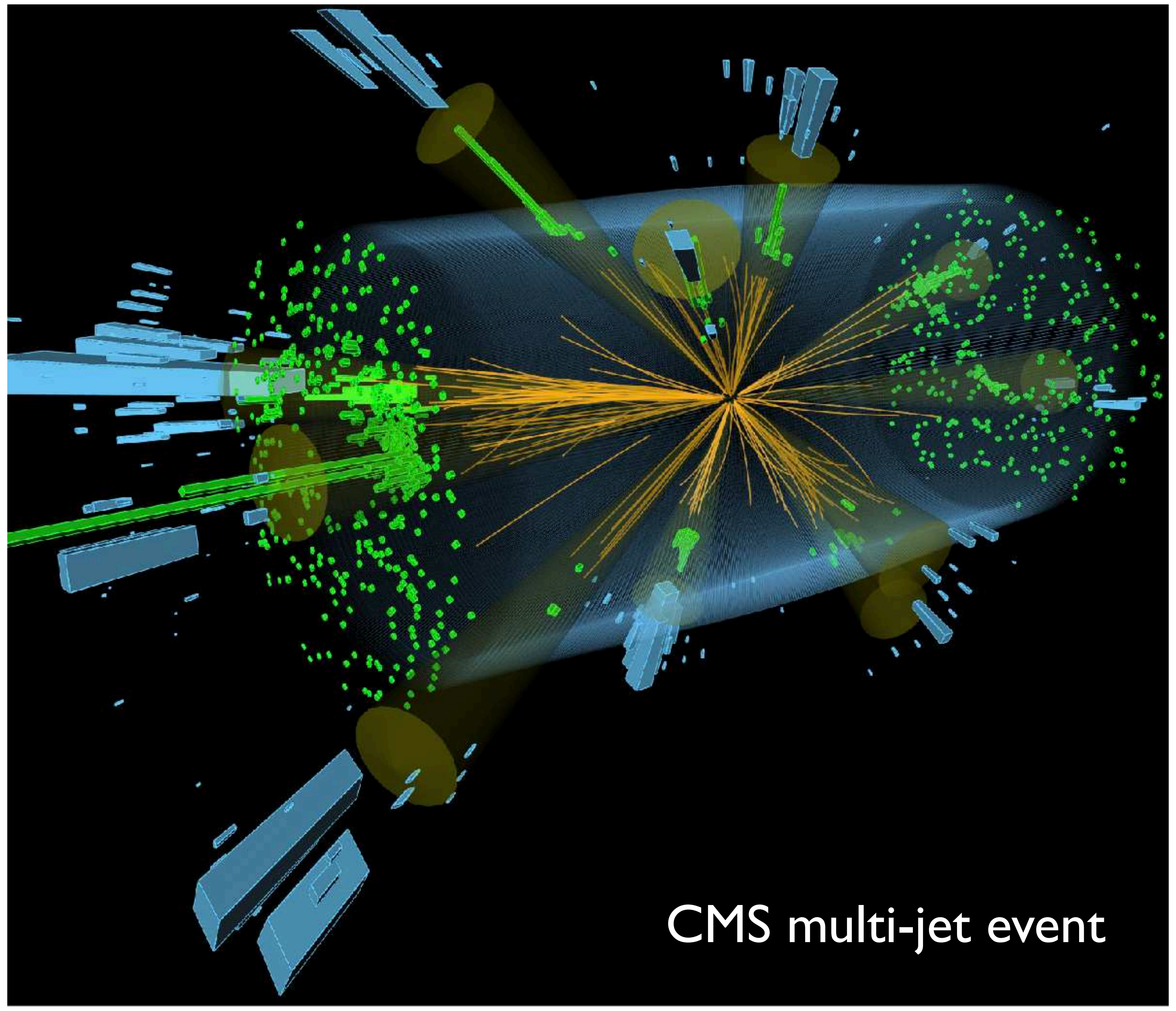
Particle physics events are naturally "point clouds"

Point cloud: "A set of data points in space" – Wikipedia



LIDAR data from self-driving car sensor

An unordered, variable length collection of particles



particle $\in \mathbb{R}^d$
$\{p_T, y, \phi, \ldots\}$

event $\in \mathbb{R}^{M\times d}$

Due to quantum-mechanical indistinguishability

Due to probabilistic nature of event formation

Deep Sets for Particle Jets

[Zaheer, Kottur, Ravanbhakhsh, Póczos, Salakhutdinov, Smola, NeurIPS 2017;
PTK, Metodiev, Thaler, JHEP 2019]

Provable decompositions of symmetric functions

(See backup for details)

Energy Flow Network (EFN)

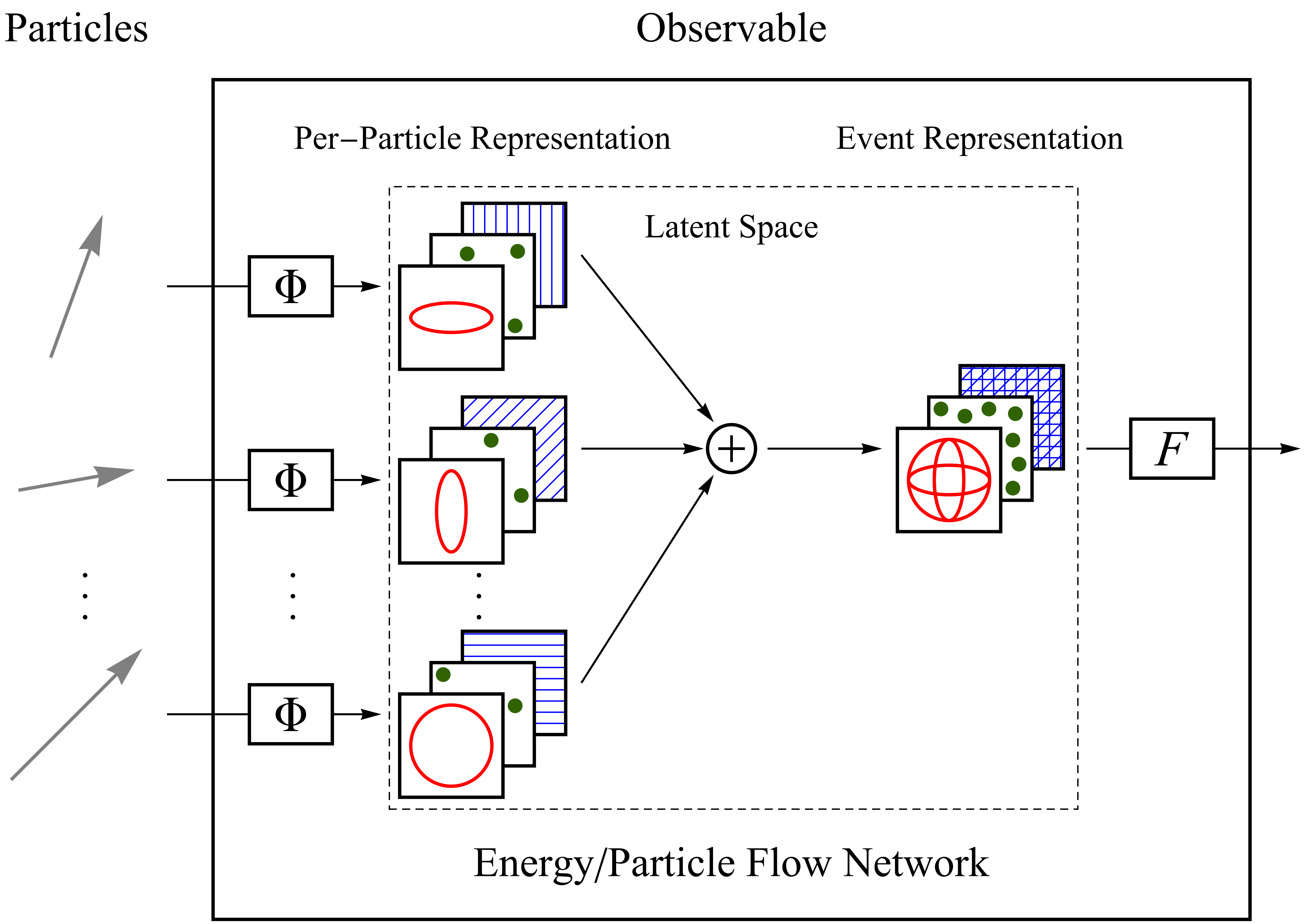
$$\mathrm{EFN}(\{p_1^\mu,\ldots,p_M^\mu\}) = F\left(\sum_{i=1}^{M} z_i \Phi(\hat{p}_i)\right)$$

Energy-weighted (IRC-safe) latent space

Particle Flow Network (PFN)

$$\mathrm{PFN}(\{p_1^\mu,\ldots,p_M^\mu\}) = F\left(\sum_{i=1}^{M} \Phi(p_i^\mu)\right)$$

Fully general latent space



EnergyFlow Python package contains EFN and PFN implementations in Tensorflow

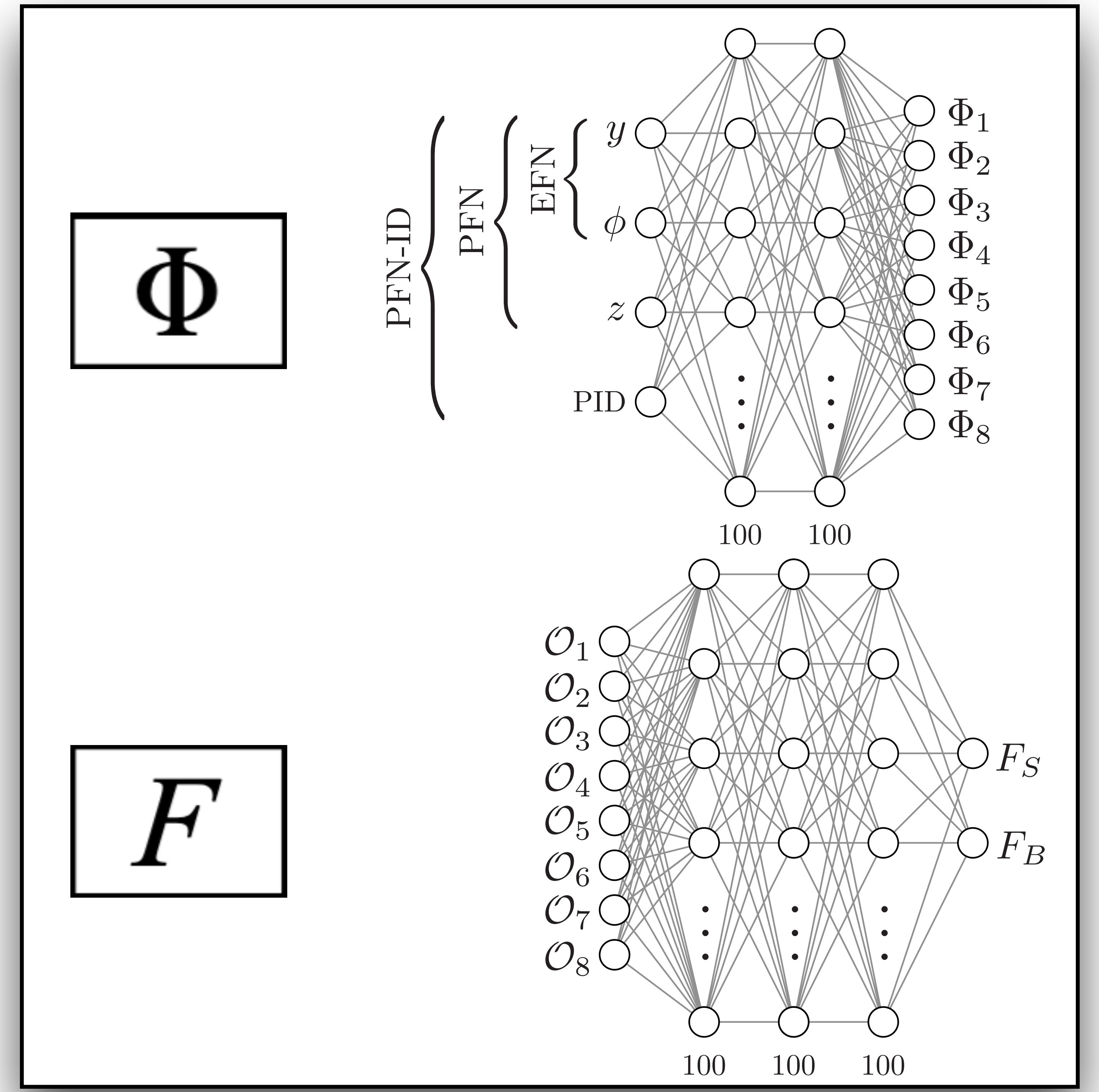
Approximating Φ and F with Neural Networks

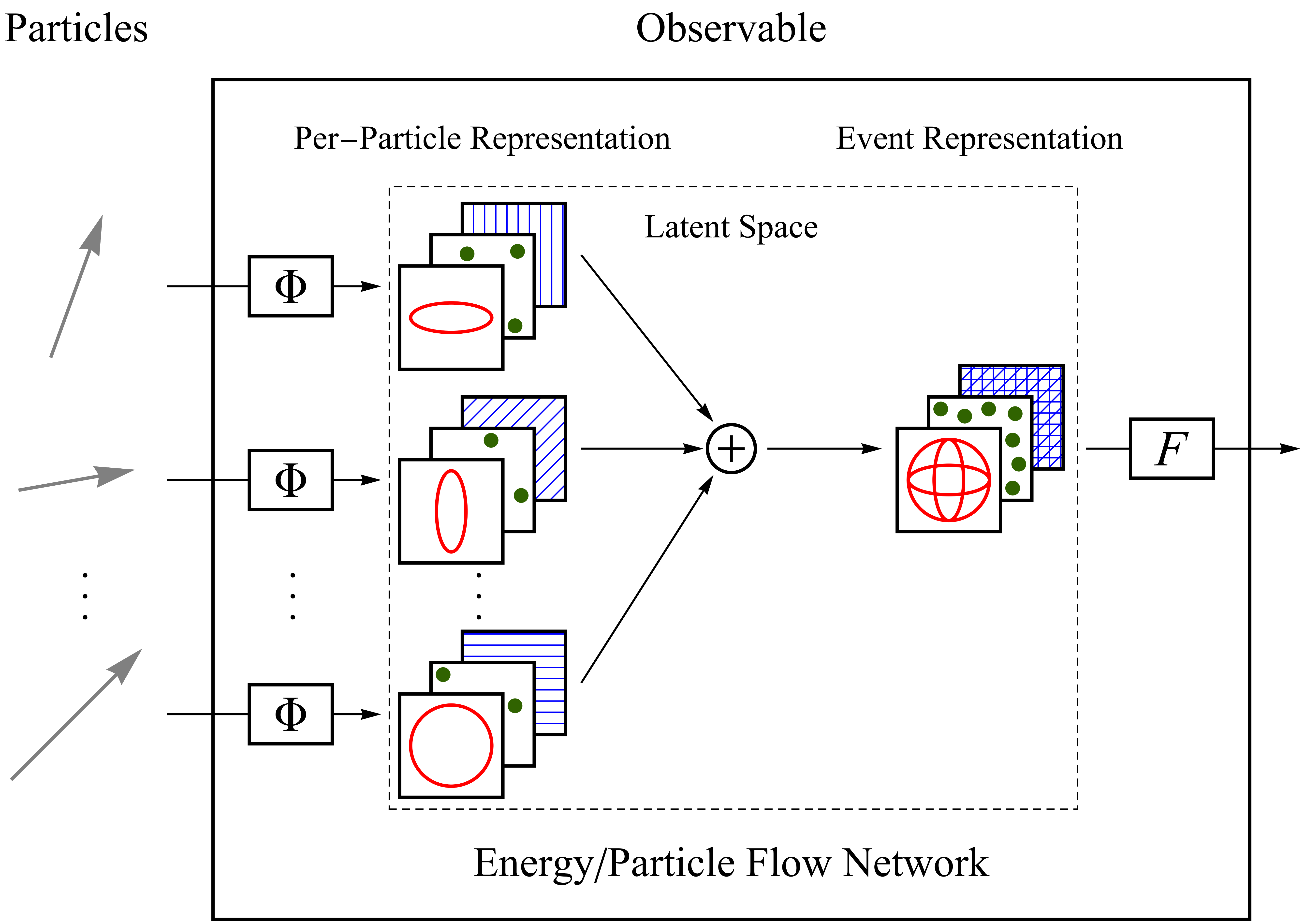
[PTK, Metodiev, Thaler, JHEP 2019]

Employ neural networks as arbitrary function approximators

Use fully-connected networks for simplicity

Default sizes – Φ: (100, 100, ℓ), F: (100, 100, 100)





EFN : $\mathcal{O}_a = \sum_{i=1}^{M} z_i \Phi_a(y_i, \phi_i)$ PFN : $\mathcal{O}_a = \sum_{i=1}^{M} \Phi_a(z_i, y_i, \phi_i, [\mathrm{PID}_i])$

Quark vs. Gluon: Latent Dimension Sweep

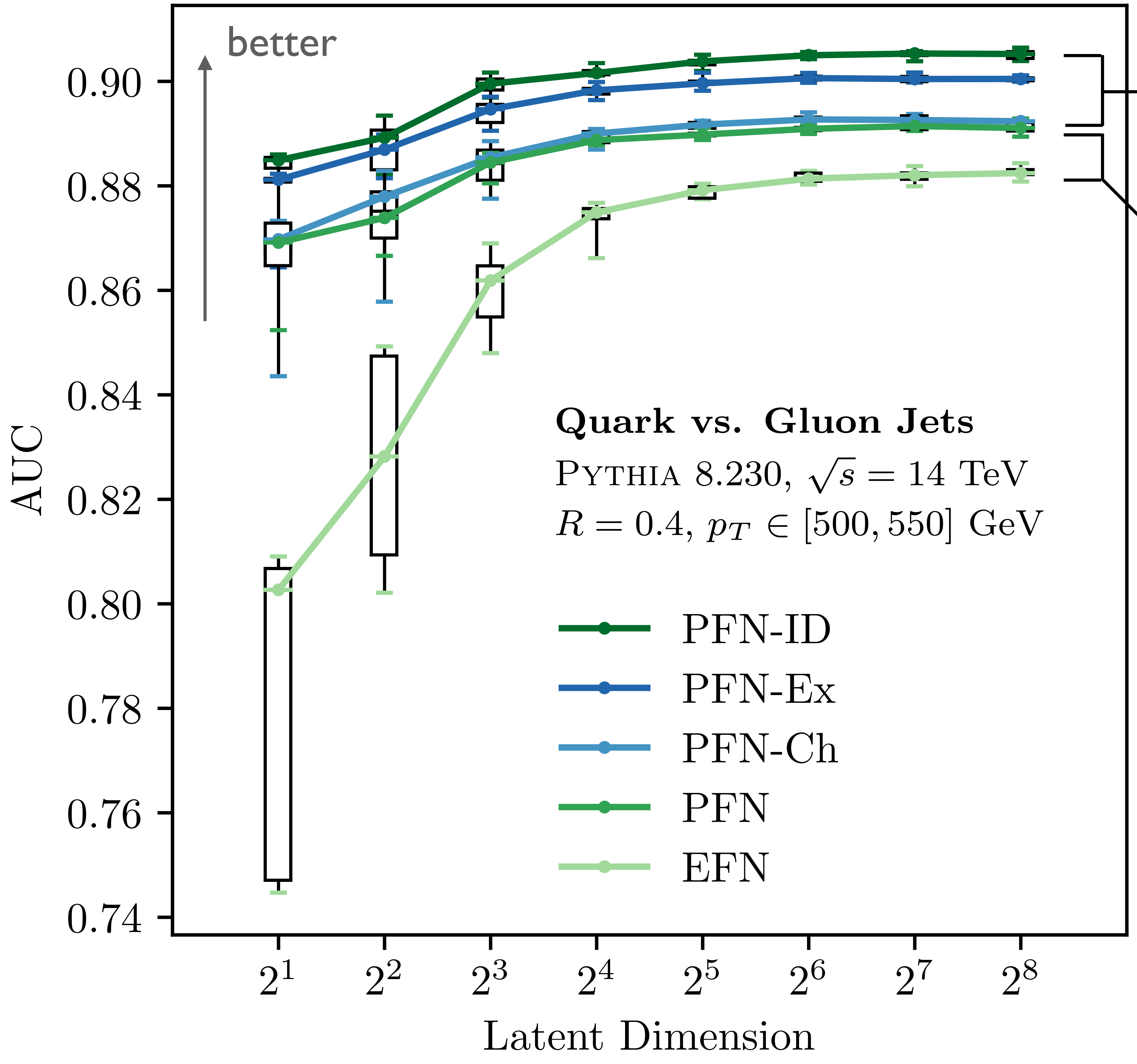
PFN-ID: Full particle flavor info
$(\gamma, \pi^\pm, K^\pm, K_L, p, \bar{p}, n, \bar{n}, e^\pm, \mu^\pm)$

PFN-Ex: Experimentally accessible info
$(\gamma, h^{\pm,0}, e^\pm, \mu^\pm)$

PFN-Ch: Particle charge info
(+,0,–)

PFN: No particle type info, arbitrary energy dependence

EFN: IRC-safe latent space



Particle type info helpful

IRC unsafe information clearly helpful

Performance saturates as dimension increases

Quark vs. Gluon: Classification Performance

[PTK, Metodiev, Thaler, JHEP 2019]

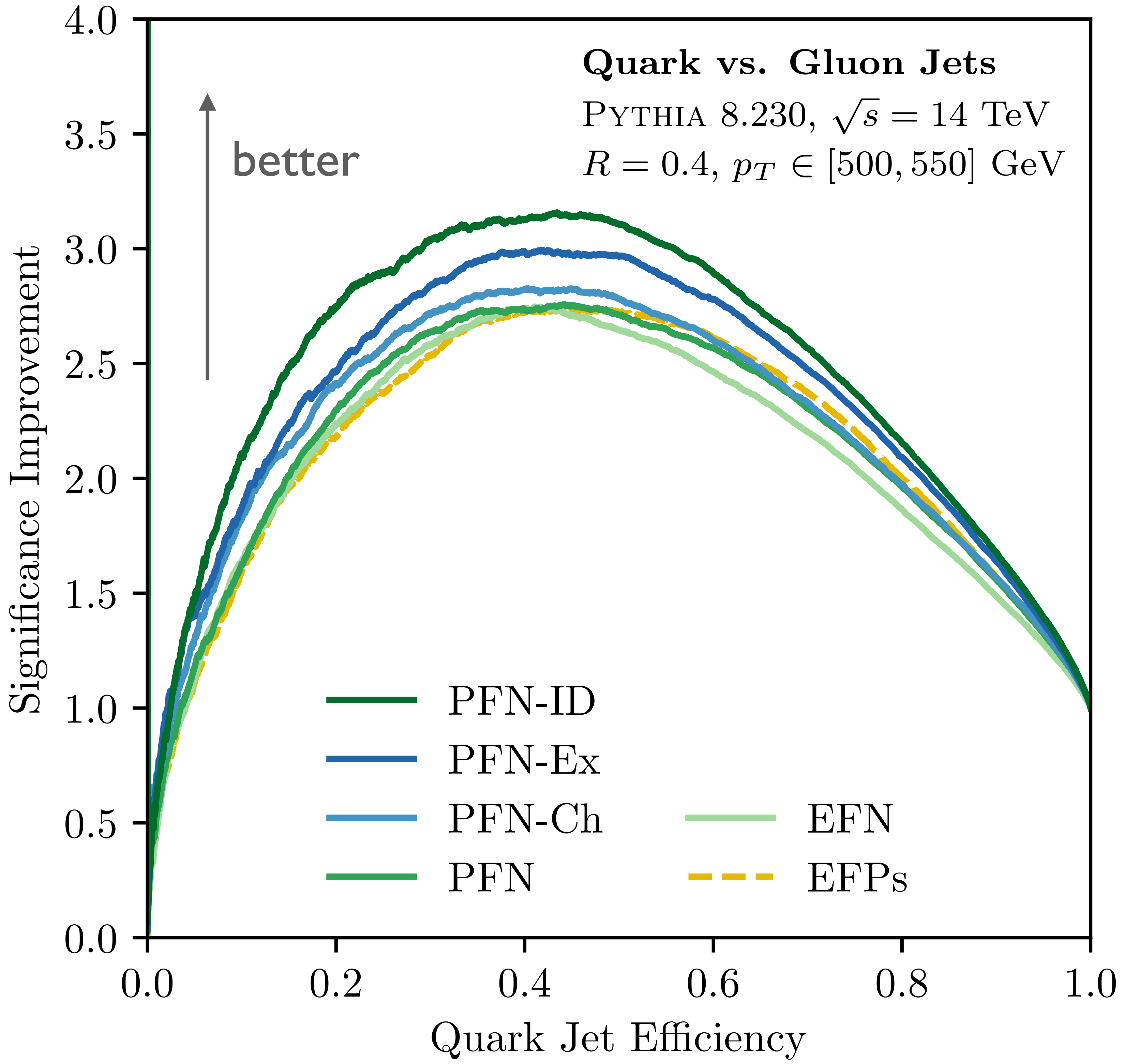
PFN-ID: Full particle flavor info
$(\gamma, \pi^\pm, K^\pm, K_L, p, \bar{p}, n, \bar{n}, e^\pm, \mu^\pm)$

PFN-Ex: Experimentally accessible info
$(\gamma, h^{\pm,0}, e^\pm, \mu^\pm)$

PFN-Ch: Particle charge info
(+,0,−)

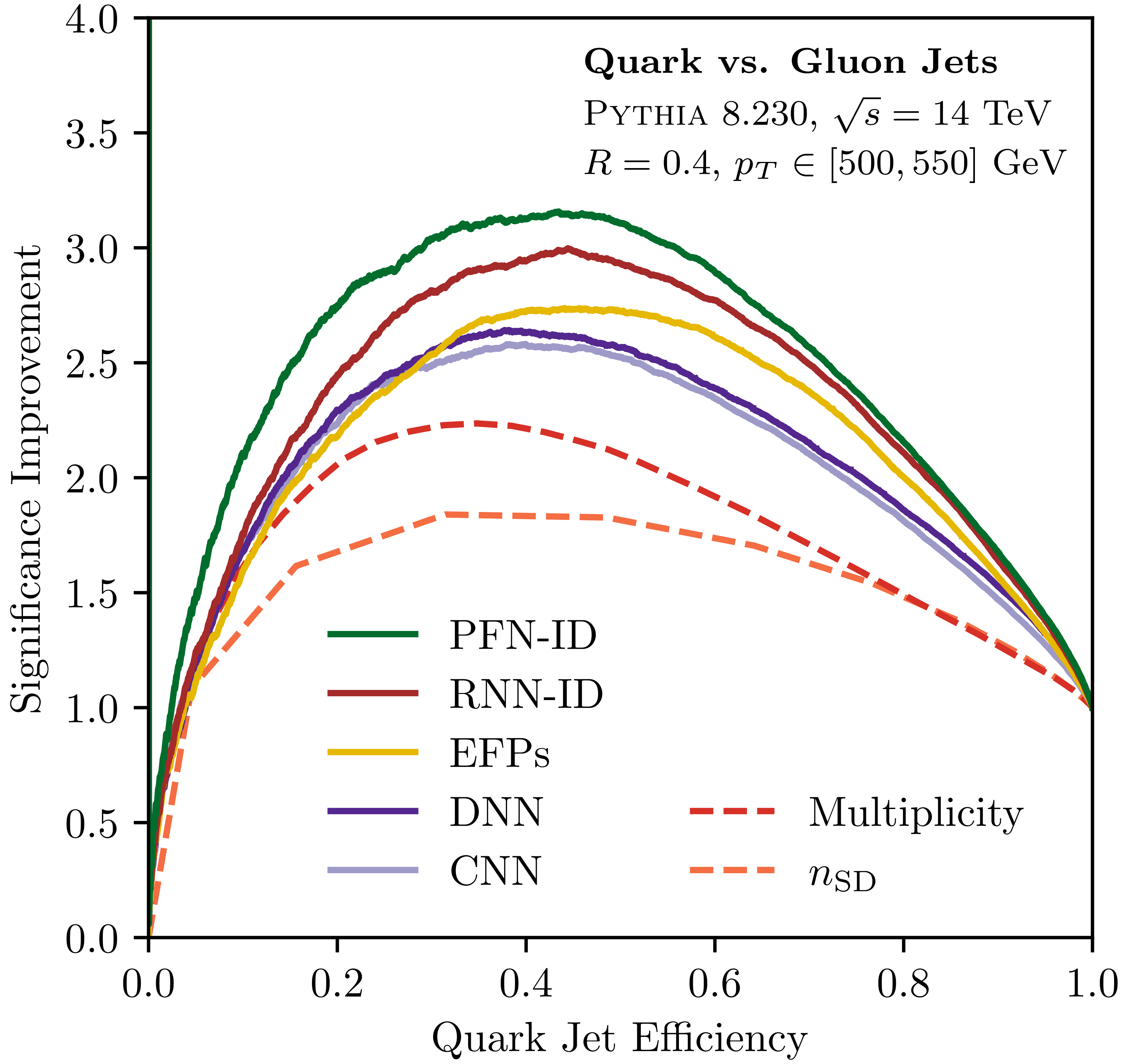
PFN: No particle type info, arbitrary energy dependence

EFN: IRC-safe latent space



Latent space dimension ℓ = 256

EFPs are comparable to EFN



PFN-ID better than RNN-ID

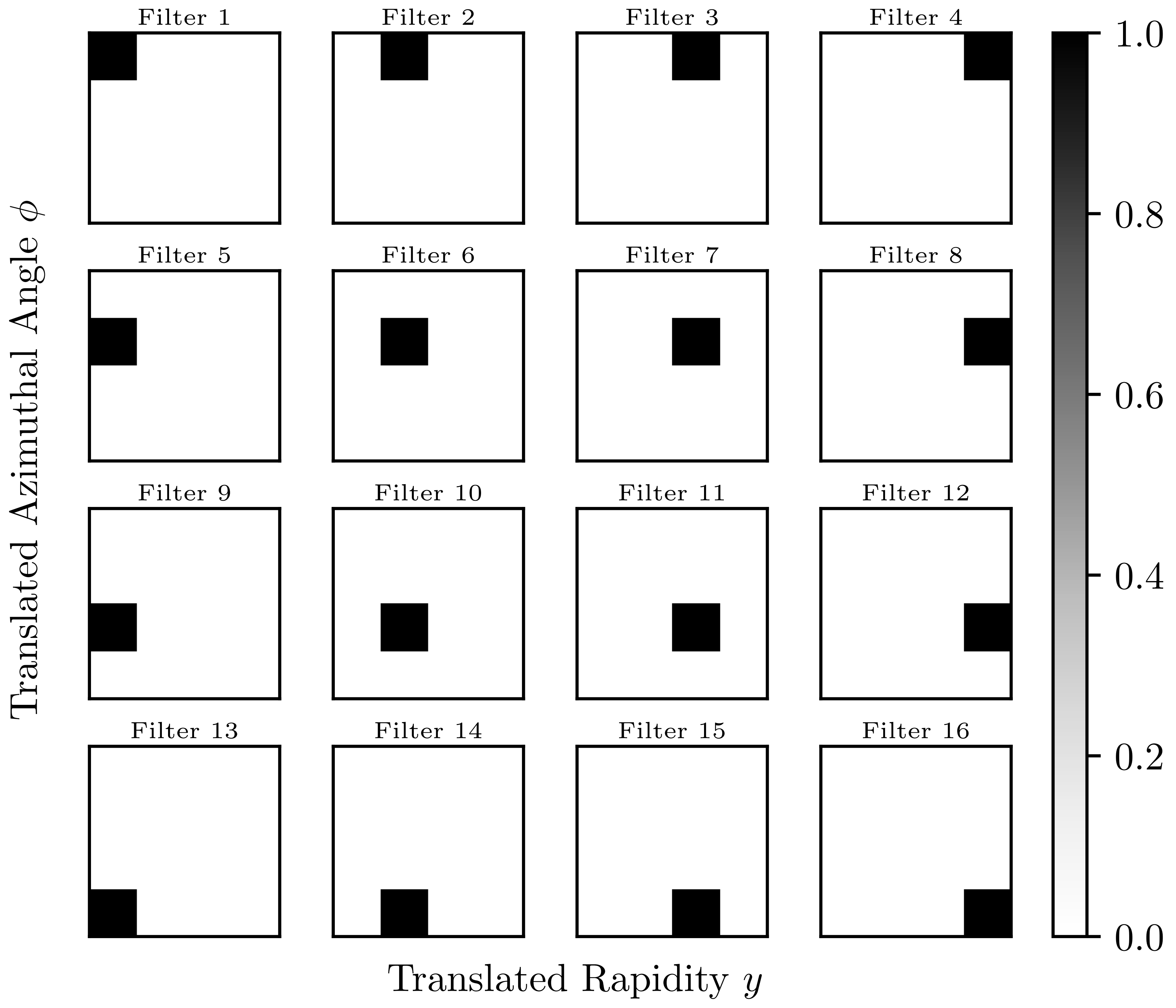
Energy Flow Network Visualization

Visualize EFN observables as 2D filters in the translated rapidity-azimuth plane

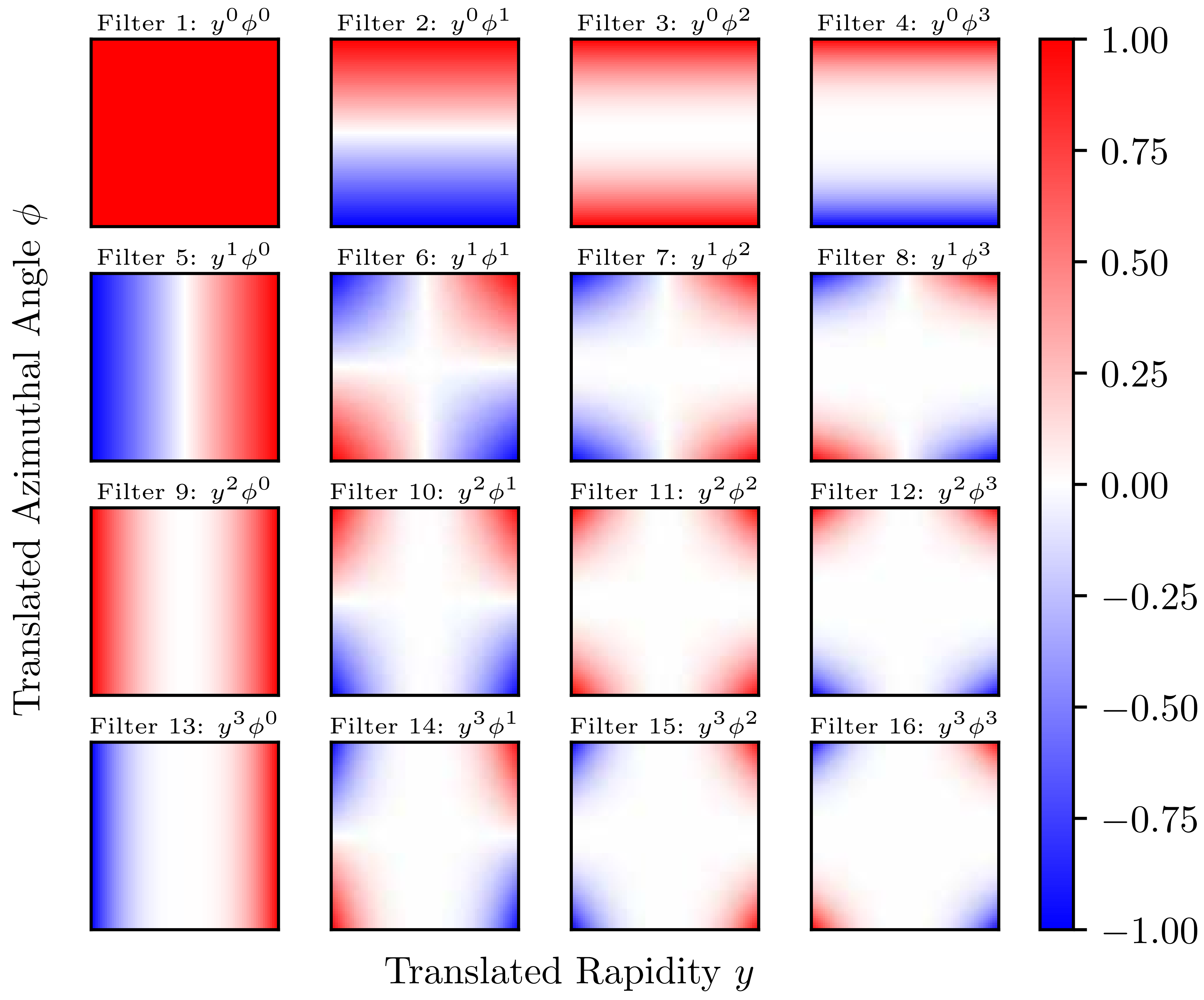
$$\text{EFN}: \ \mathcal{O}_a = \sum_{i=1}^{M} z_i \Phi_a(y_i, \phi_i)$$

Two-Dimensional Position

Jet images as EFN filters



Moments as EFN filters

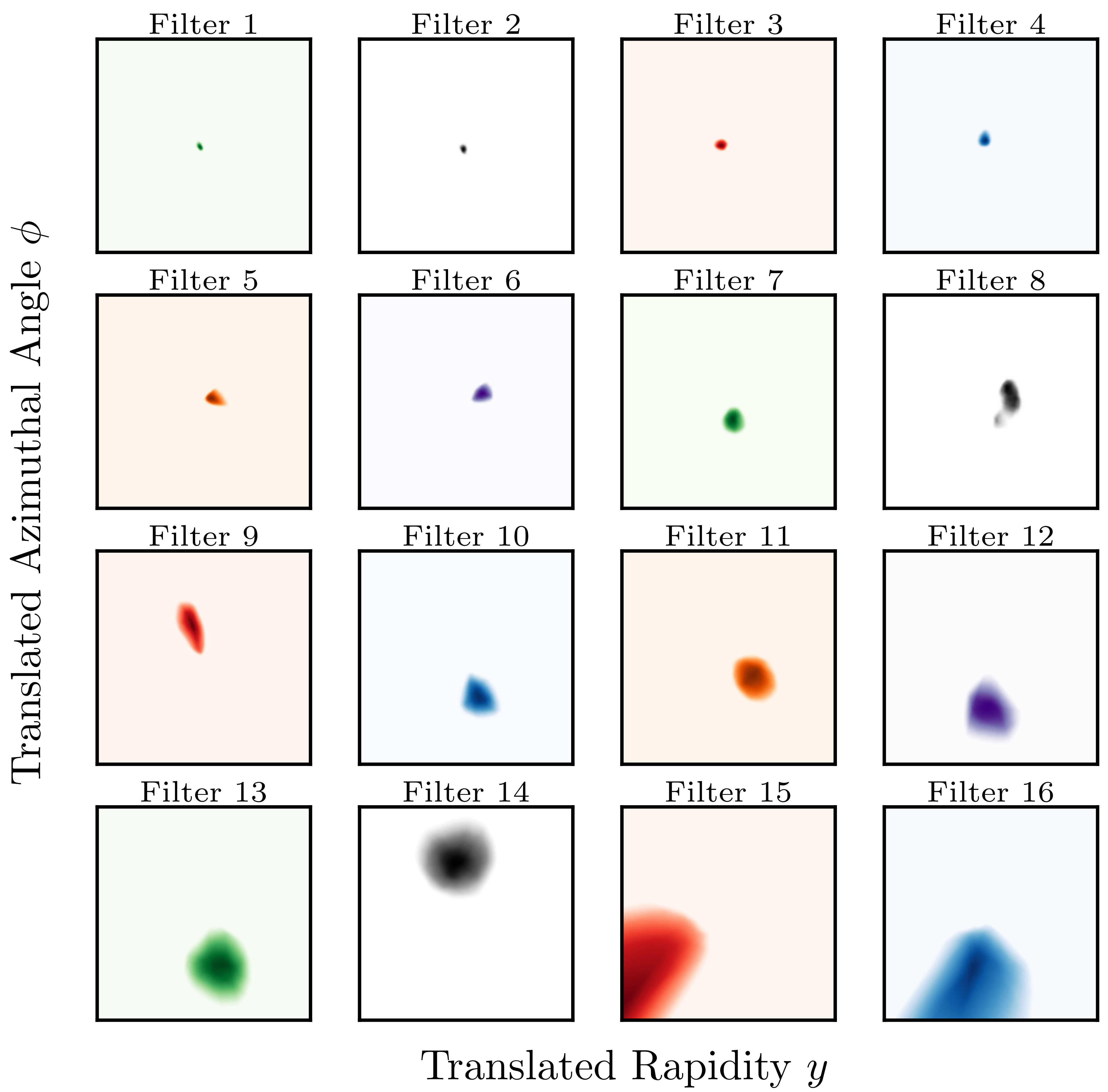


[Donoghue, Low, Pi, PRD 1979; Gur-Ari, Papucci, Perez, 1101.2905; PTK, Metodiev, Thaler, PRD 2020]

Energy Flow Network Visualization – Quark vs. Gluon

EFN (ℓ = 256) randomly selected filters, sorted by size

[PTK, Metodiev, Thaler, JHEP 2019]



$$\text{EFN}: \quad \mathcal{O}_a = \sum_{i=1}^{M} z_i \Phi_a(y_i, \phi_i)$$

Two-Dimensional Position

Generally find blobs of all scales

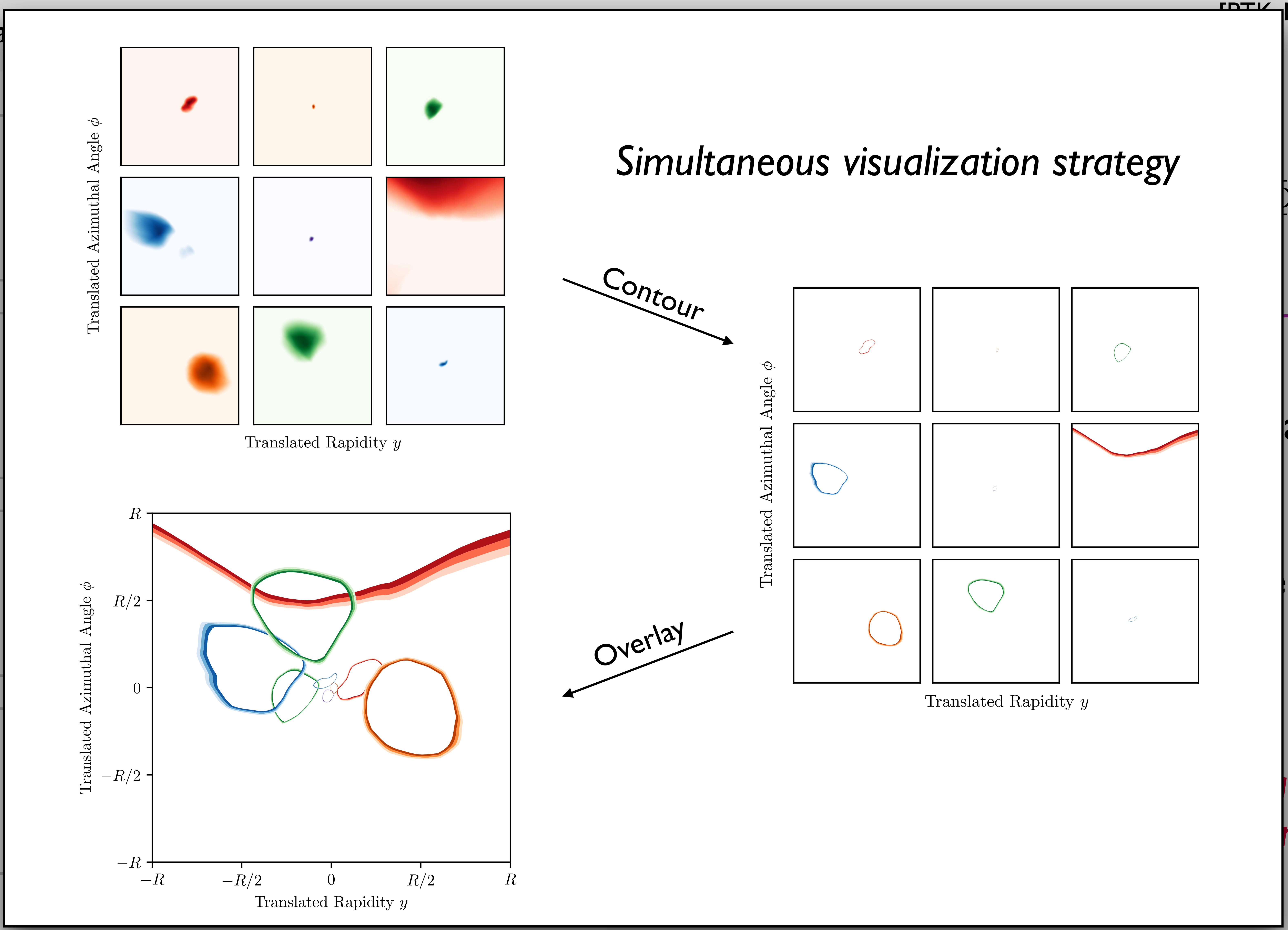
Local nature of activated region lends interpretation as "pixels"

EFN seems to have learned a dynamically sized jet image!

Quark vs. Gluon: Visualizing EFN Filters

EFN (ℓ = 256) ra

[PTK, Metodiev, Thaler, JHEP 2019]



$\Phi_a(y_i, \phi_i)$

-Dimensional Position

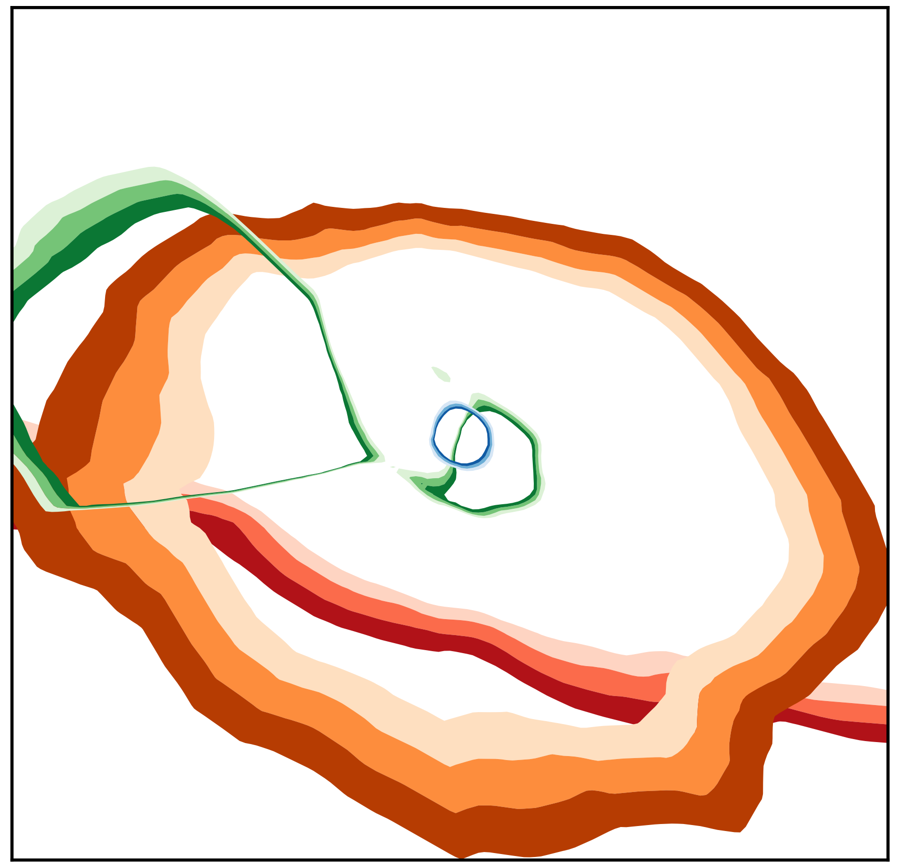
all scales

d region
"pixels"

rned a
mage!

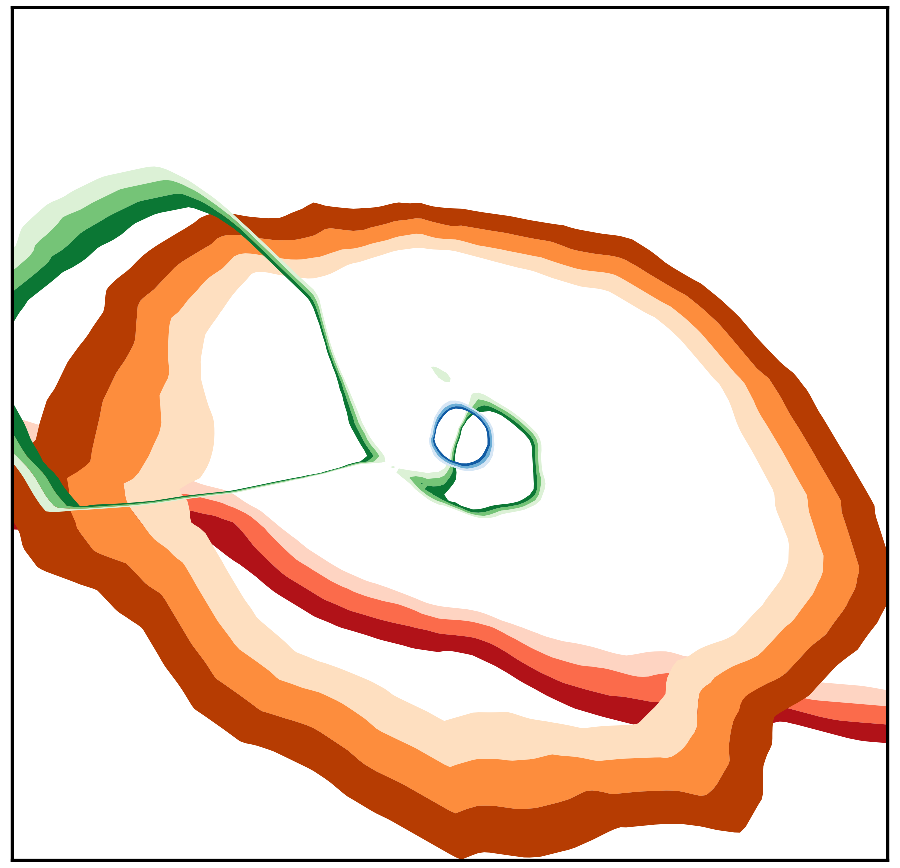
Quark vs. Gluon: Visualizing EFN Filters

Quark vs. Gluon: Visualizing EFN Filters

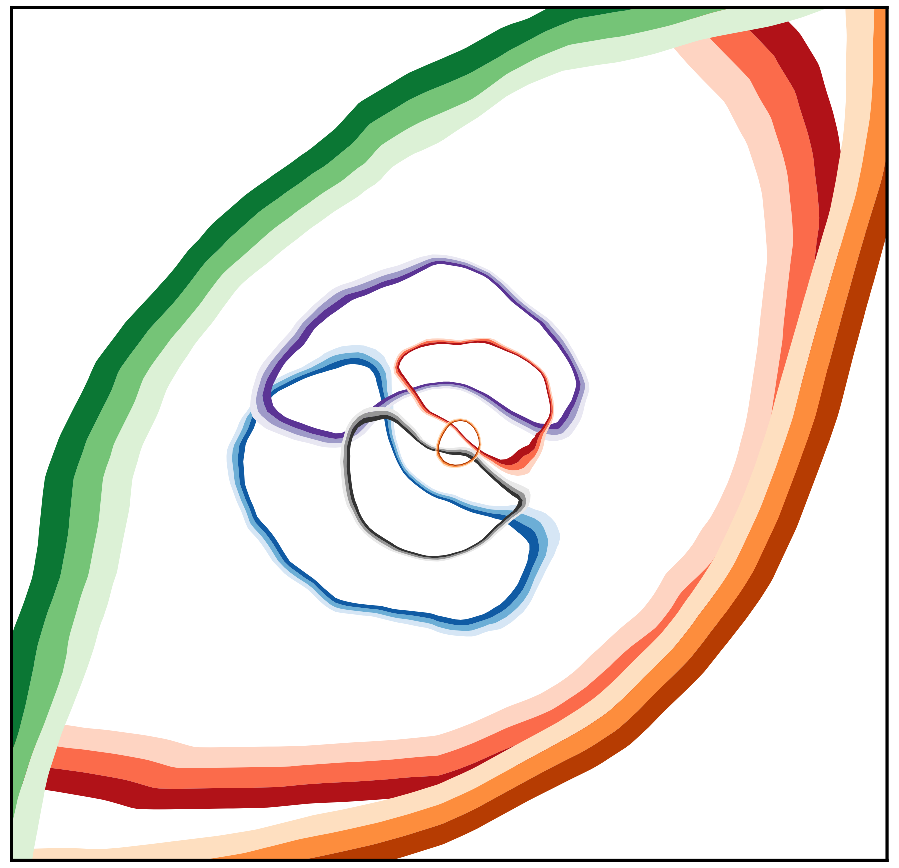


$\ell = 4$

Quark vs. Gluon: Visualizing EFN Filters

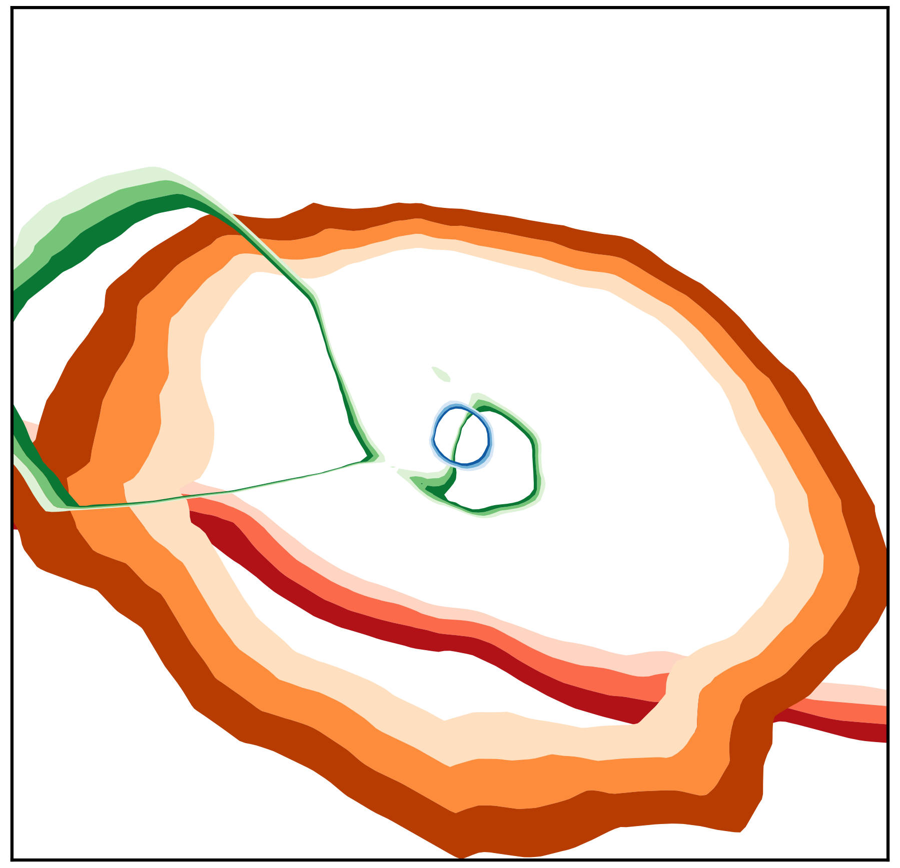


$\ell = 4$

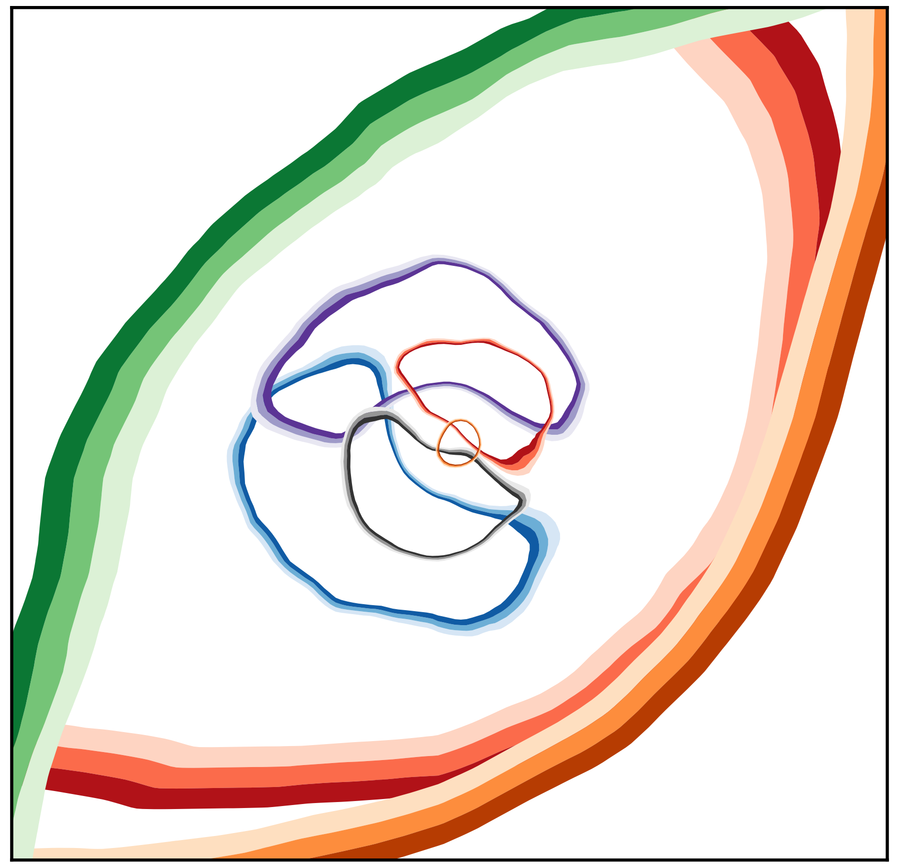


$\ell = 8$

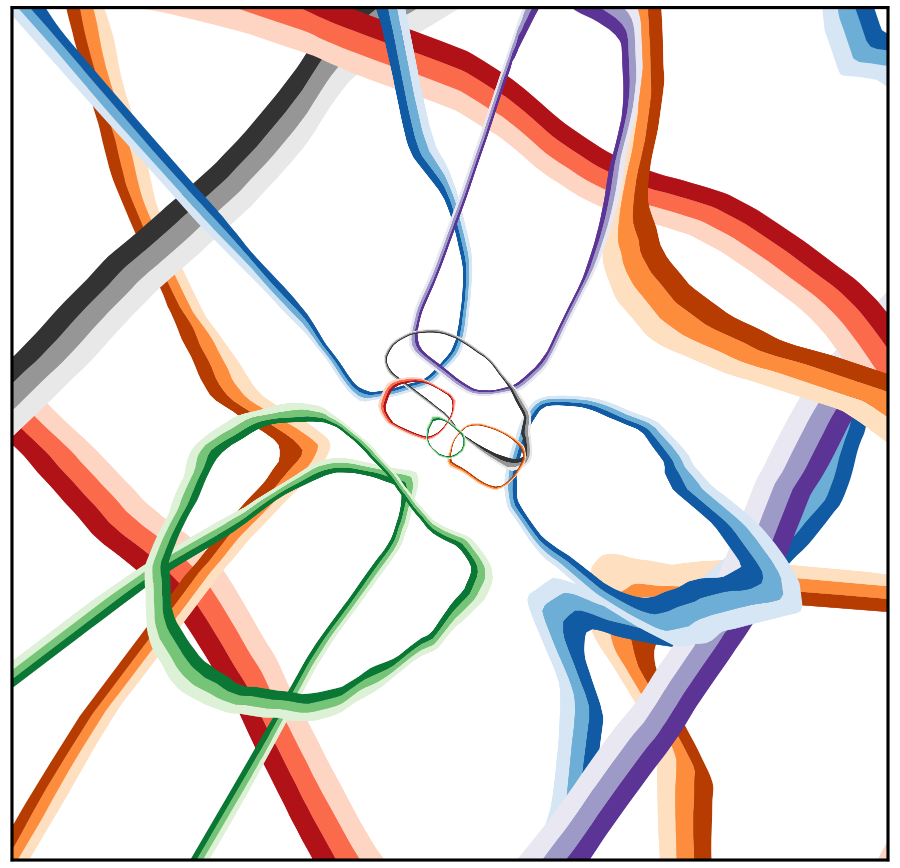
Quark vs. Gluon: Visualizing EFN Filters



$\ell = 4$

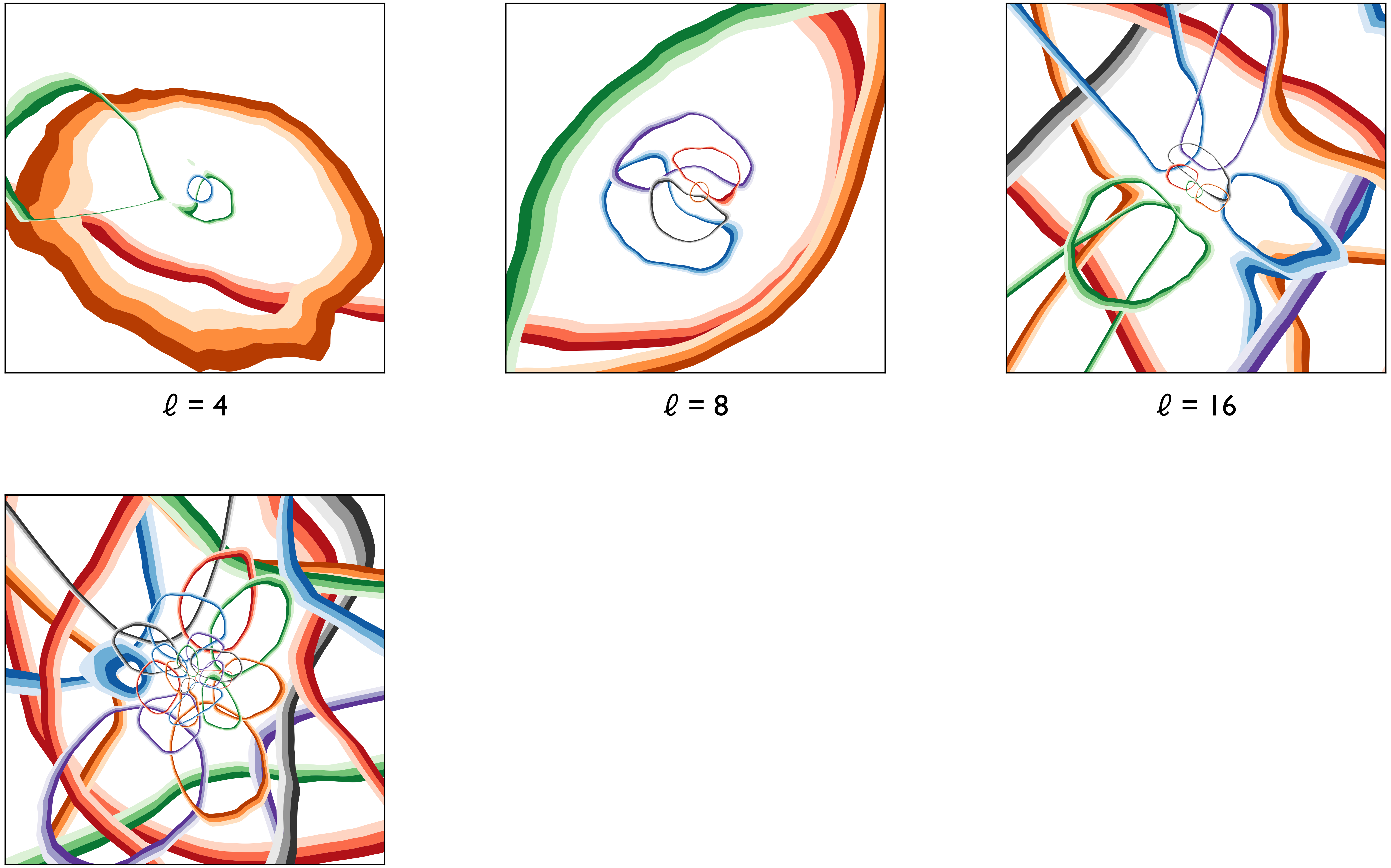


$\ell = 8$



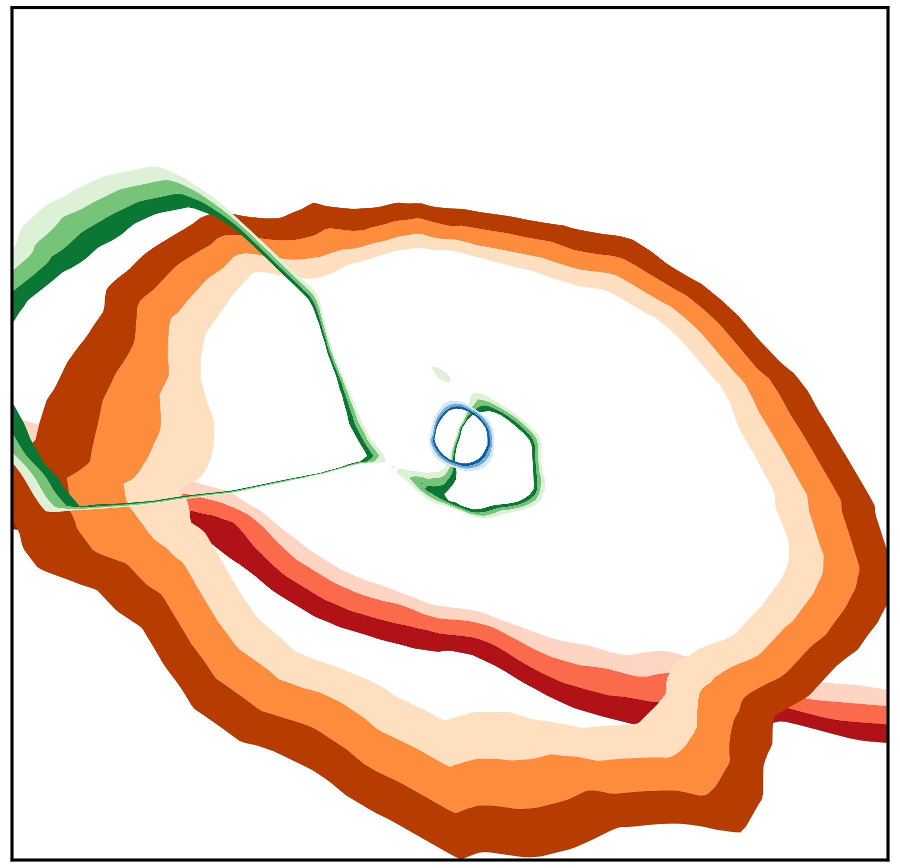
$\ell = 16$

Quark vs. Gluon: Visualizing EFN Filters

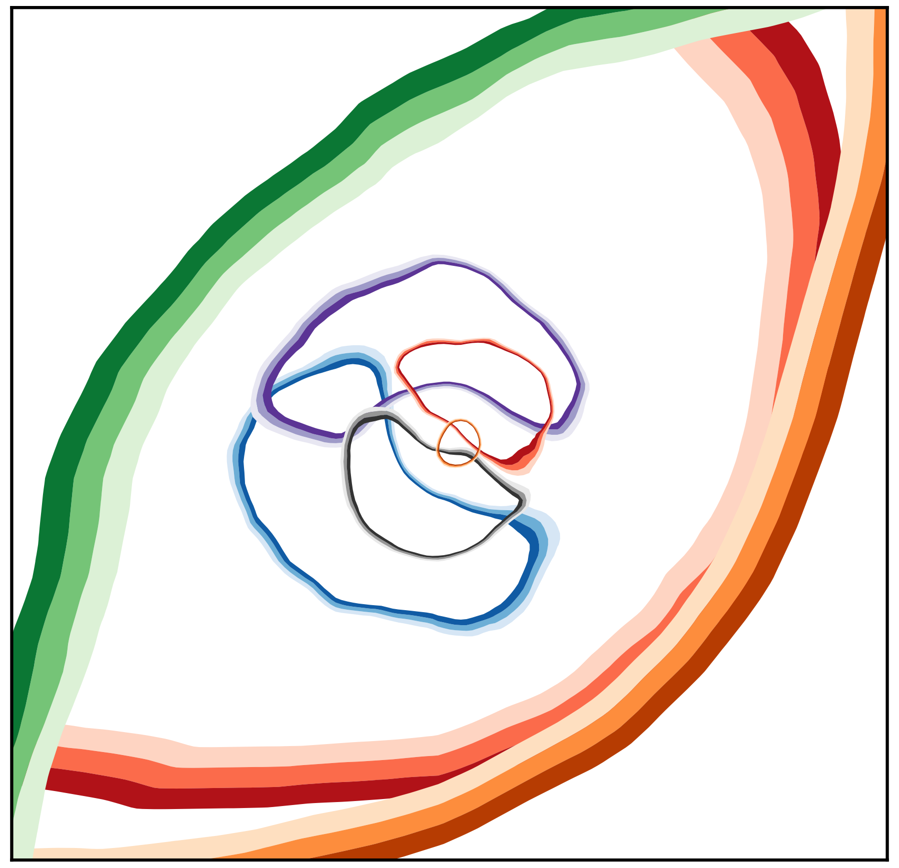


$\ell = 4$ $\ell = 8$ $\ell = 16$ $\ell = 32$

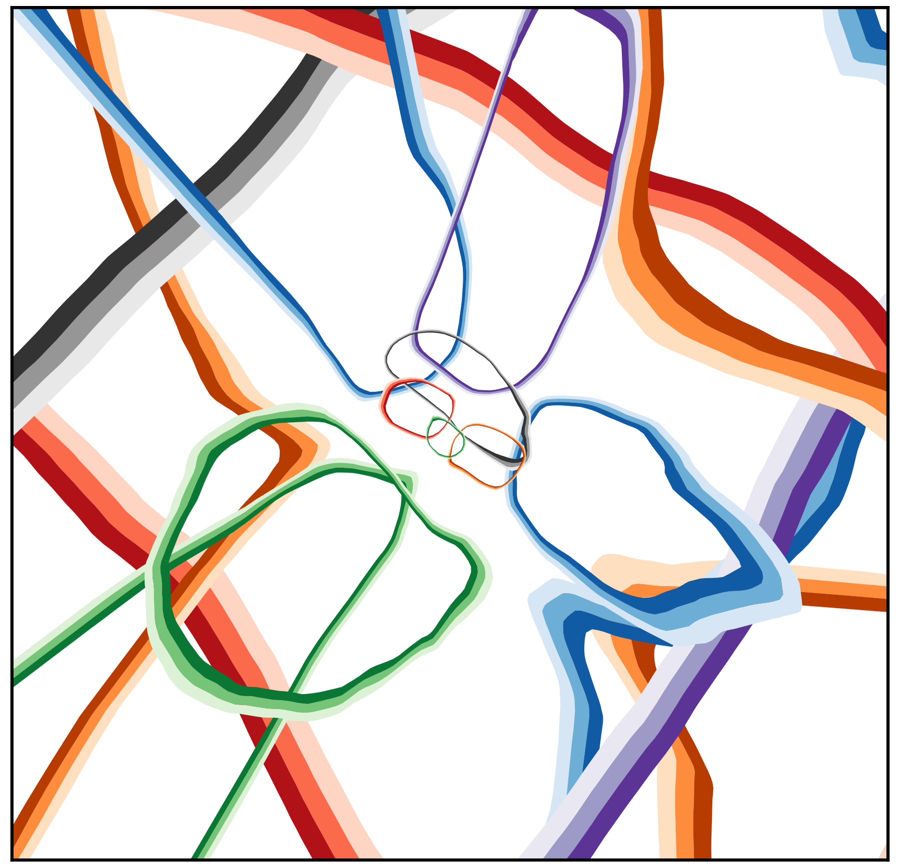
Quark vs. Gluon: Visualizing EFN Filters



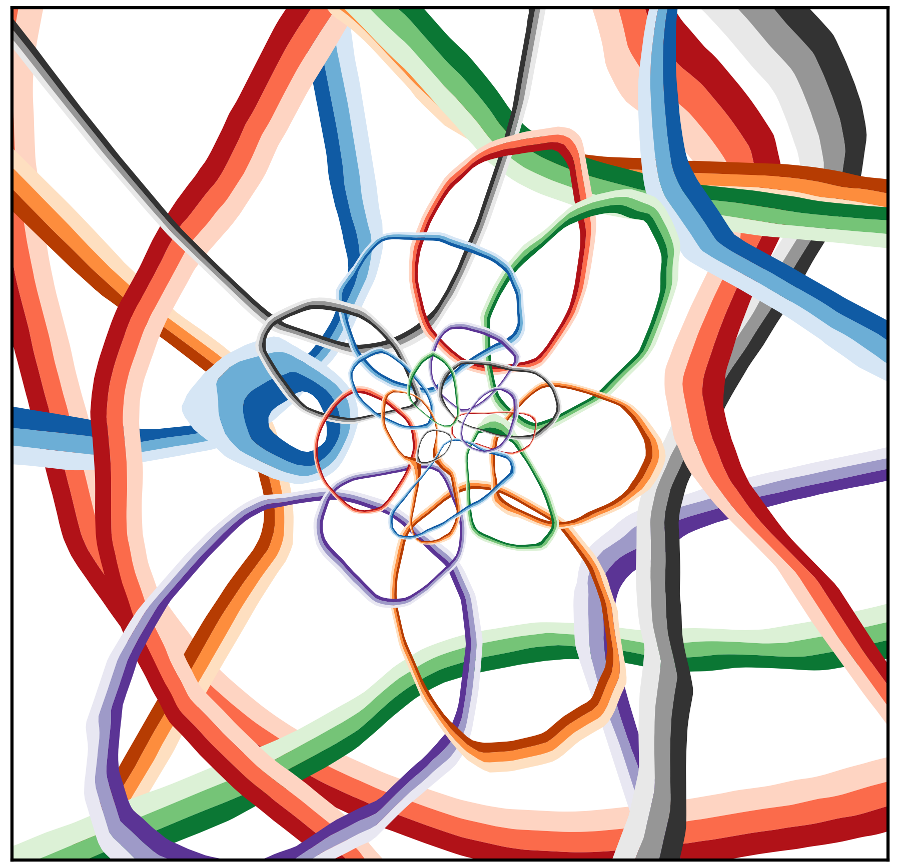
$\ell = 4$



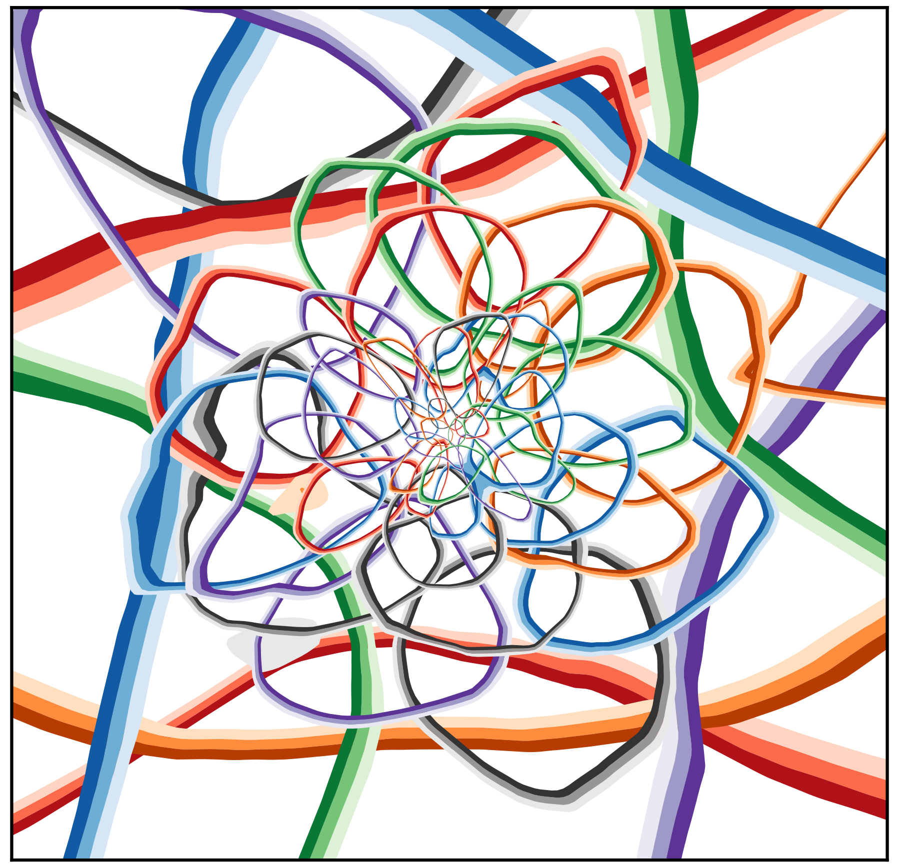
$\ell = 8$



$\ell = 16$

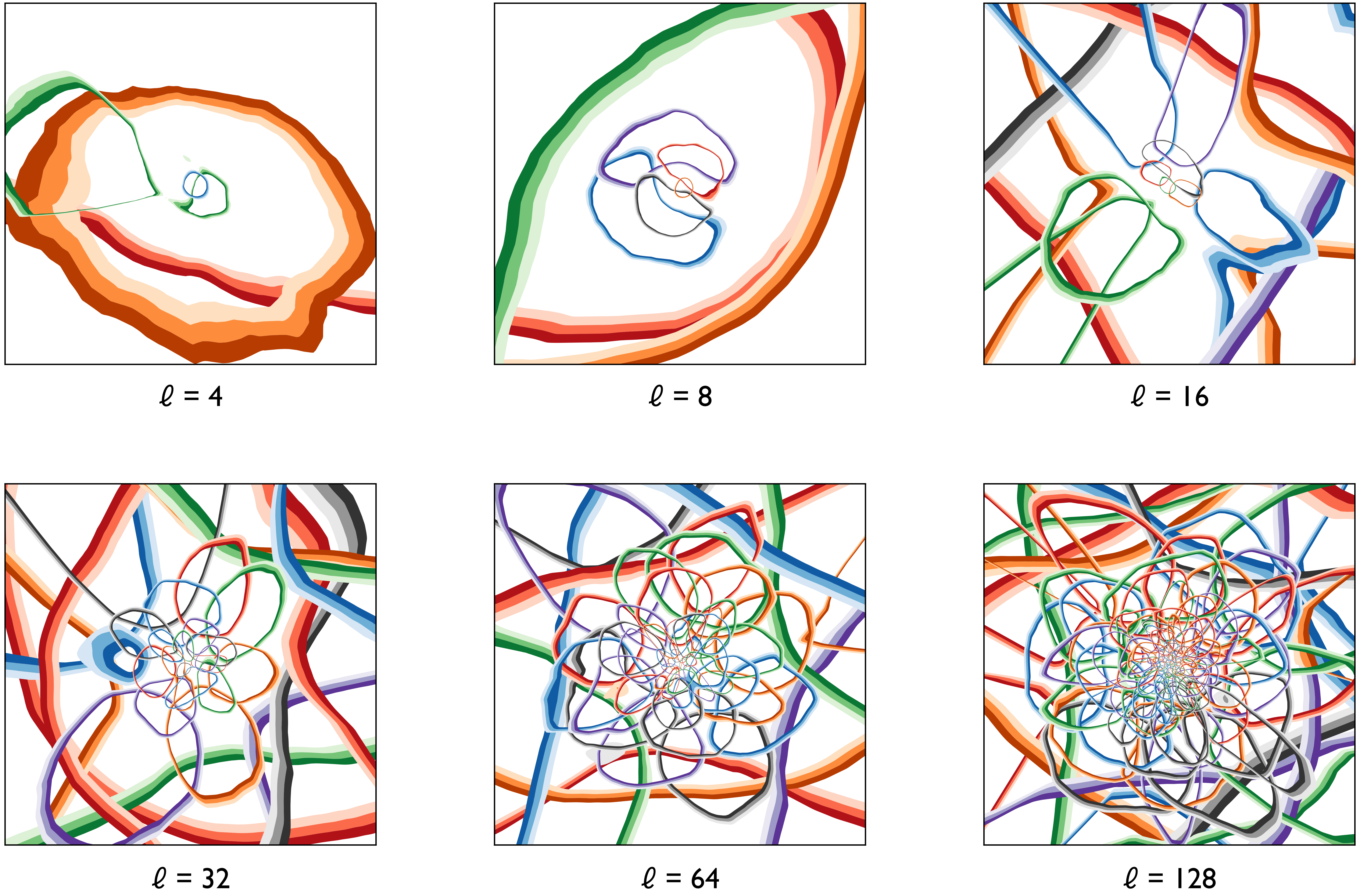


$\ell = 32$

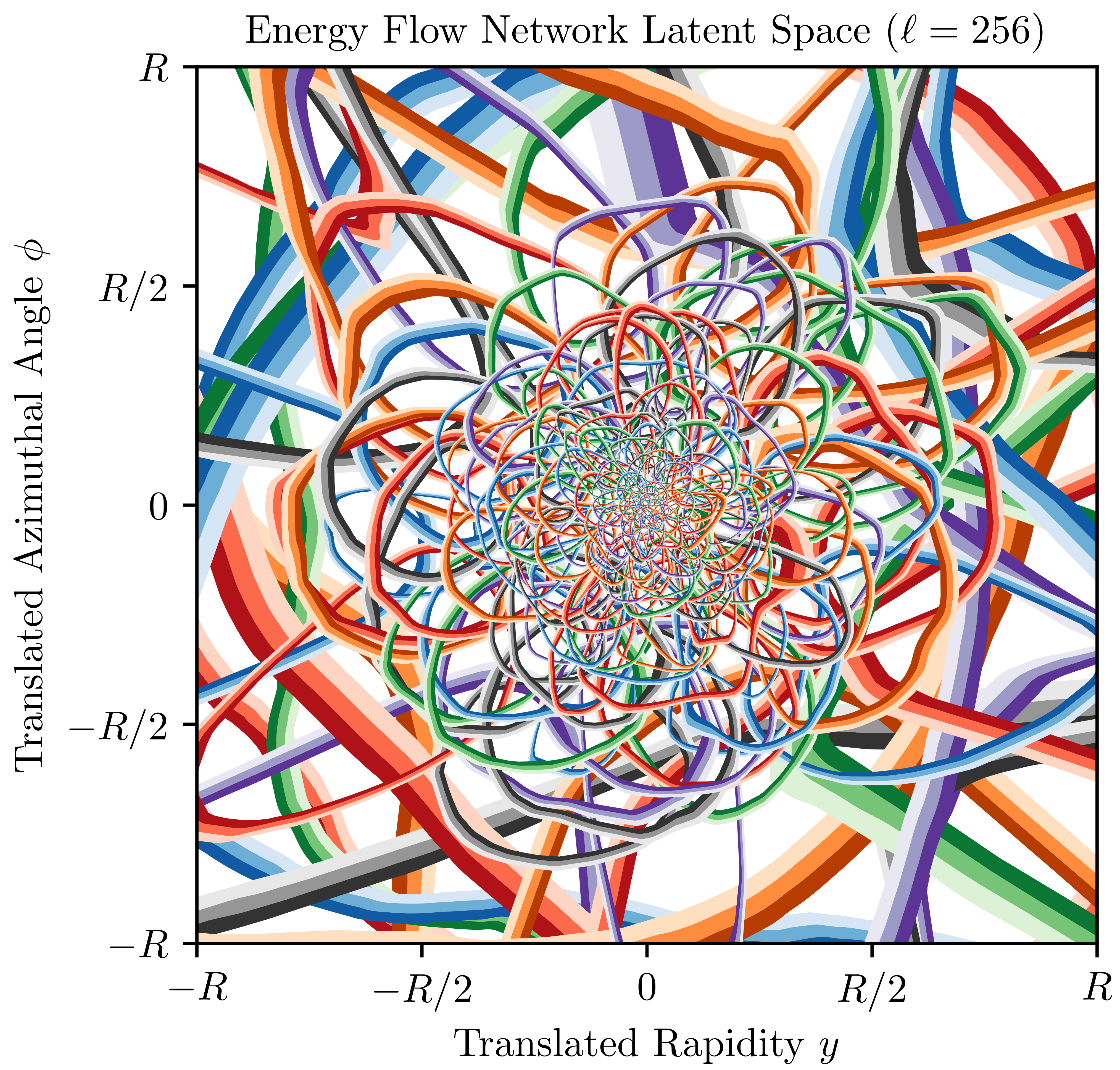


$\ell = 64$

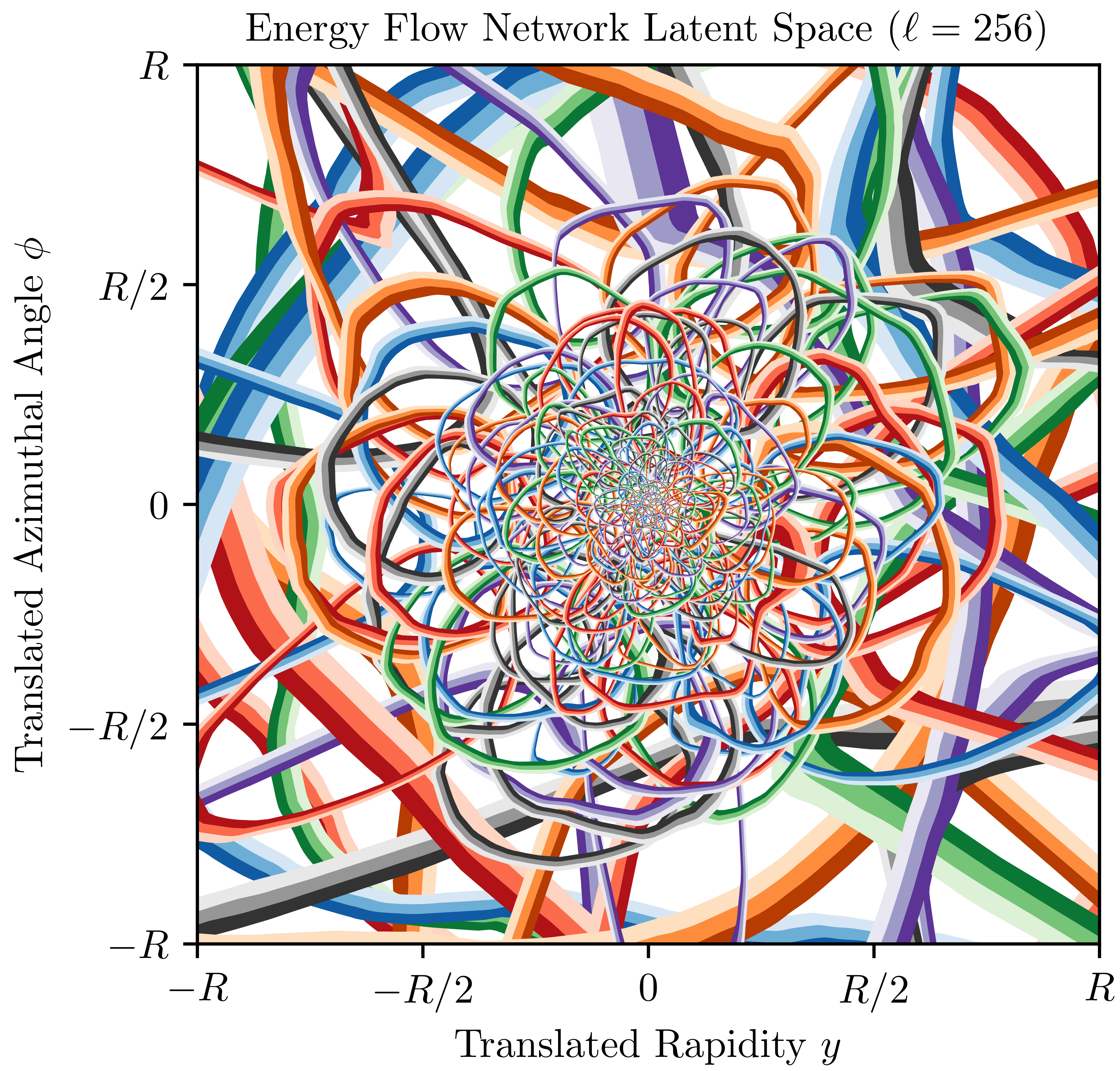
Quark vs. Gluon: Visualizing EFN Filters



Quark vs. Gluon: Visualizing EFN Filters



Quark vs. Gluon: Visualizing EFN Filters



The New York Times

Image: Alex Eben Meyer

"It looked like a pile of multicolored rubber bands, but it represented several layers of processing […]"

– Dennis Overbye, Can a Computer Devise a Theory of Everything? *New York Times*, November 23, 2020.

Quark vs. Gluon: Measuring EFN Filters

Energy Flow Network appears to have learned about the collinear singularity of QCD!

Slope of 2 is predicted by QCD

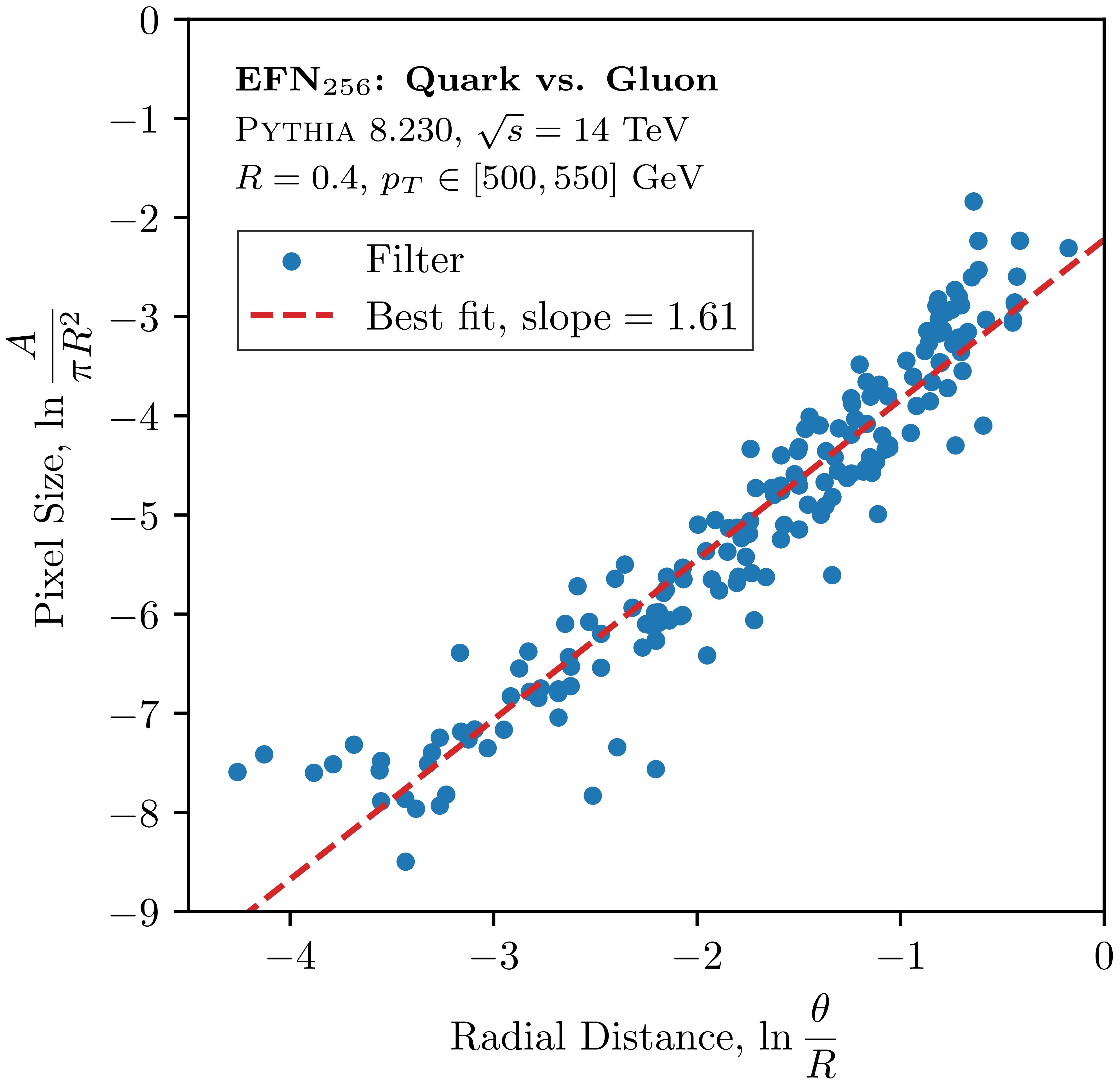
$$\left\| \mathrm{d}\ln\frac{\theta}{R}\,\mathrm{d}\varphi \right\| = \theta^2 \,\mathrm{d}y\,\mathrm{d}\phi$$

Area element in rap-phi plane

Emission plane area element

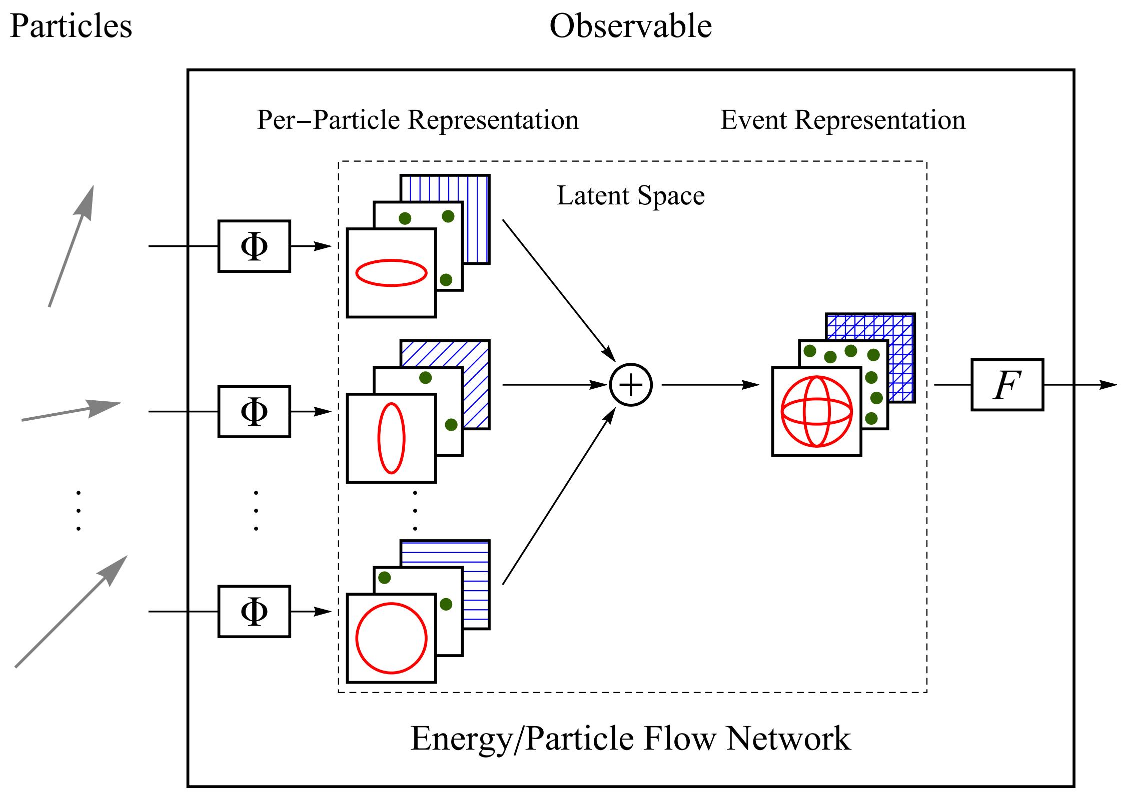
$$dP_{i\to ig} \simeq \frac{2\alpha_s}{\pi} C_a \frac{d\theta}{\theta}\frac{dz}{z}$$

Altarelli-Parisi splitting function describes gluon emission



Power-law dependence between filter size and distance from center is observed

(Additional analysis of filters in backup)



Particle Physics Fundamentals – Jets

Architectures for Colliders – EFNs/PFNs

Simple, extensible neural network architecture(s) for collider events

Statistical Deconvolution

Particle Physics Fundamentals – Jets

Jets are critical to the success of the modern collider program

Architectures for Colliders – EFNs/PFNs

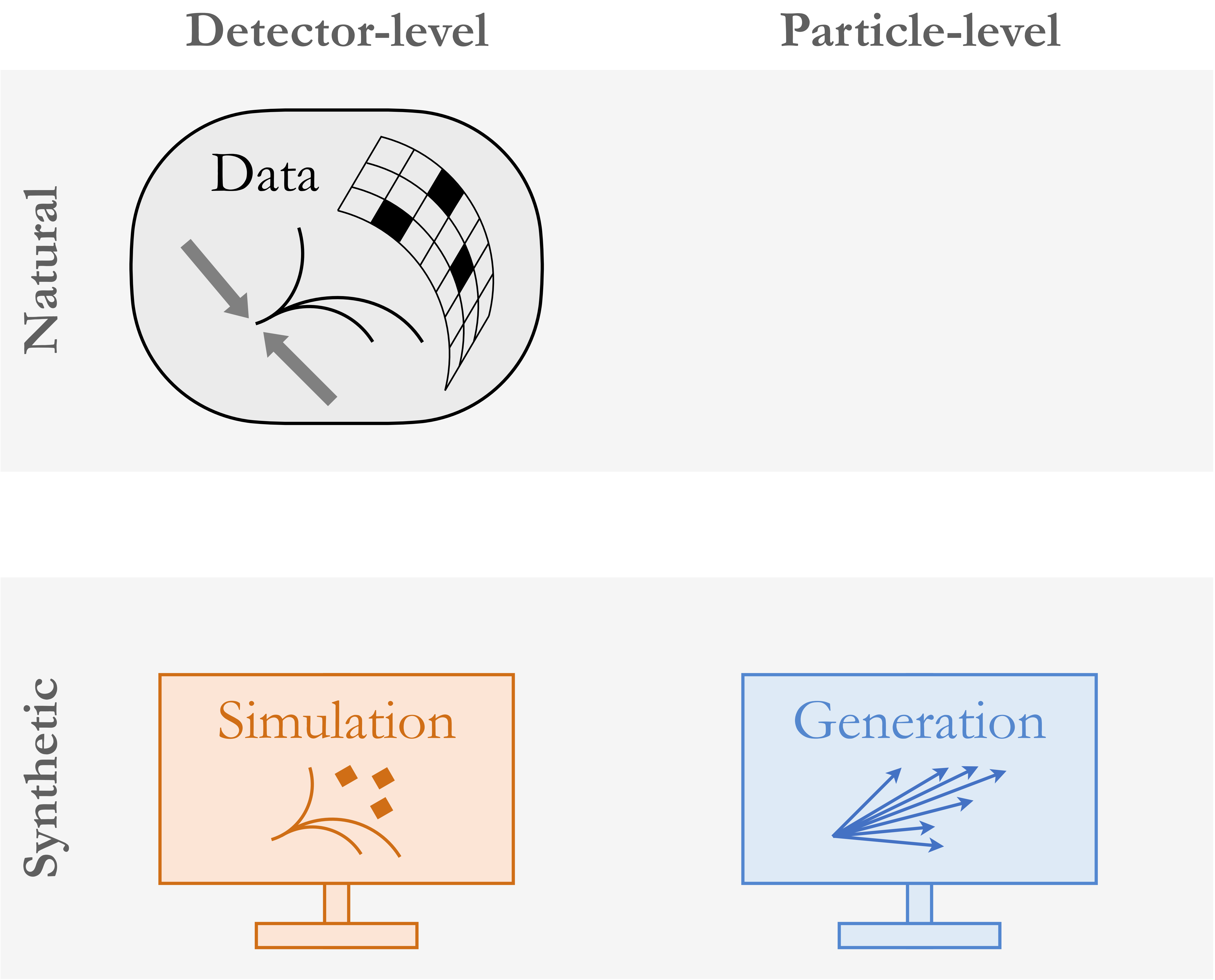
Simple, extensible neural network architecture(s) for collider events

Statistical Deconvolution

Can ML overcome the curse of dimensionality in correcting mis-measurements?

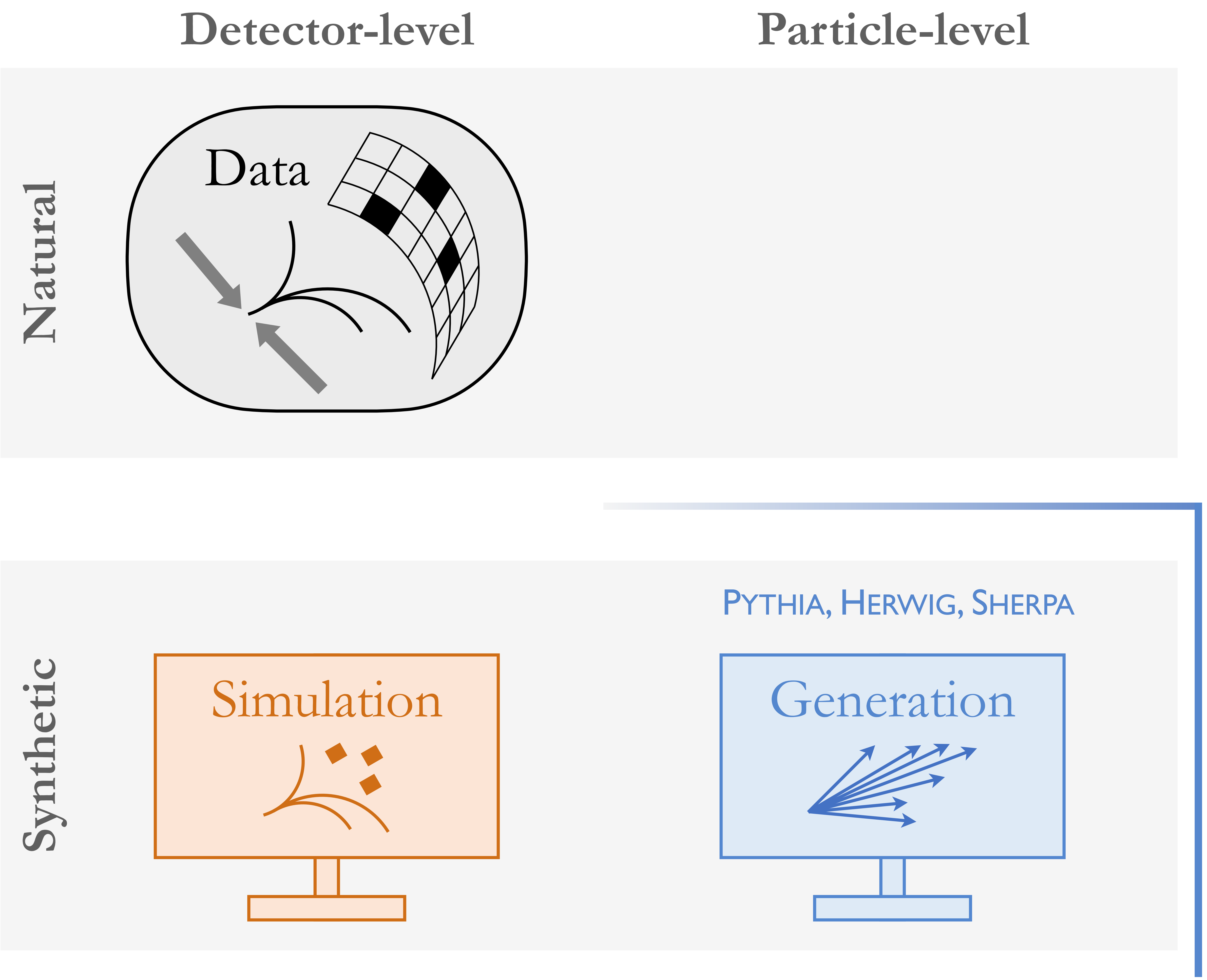
Correcting for Detector Effects – a.k.a. "Unfolding"

Measurements are affected by *detector effects* such as finite resolution, miscalibration, and limited acceptance



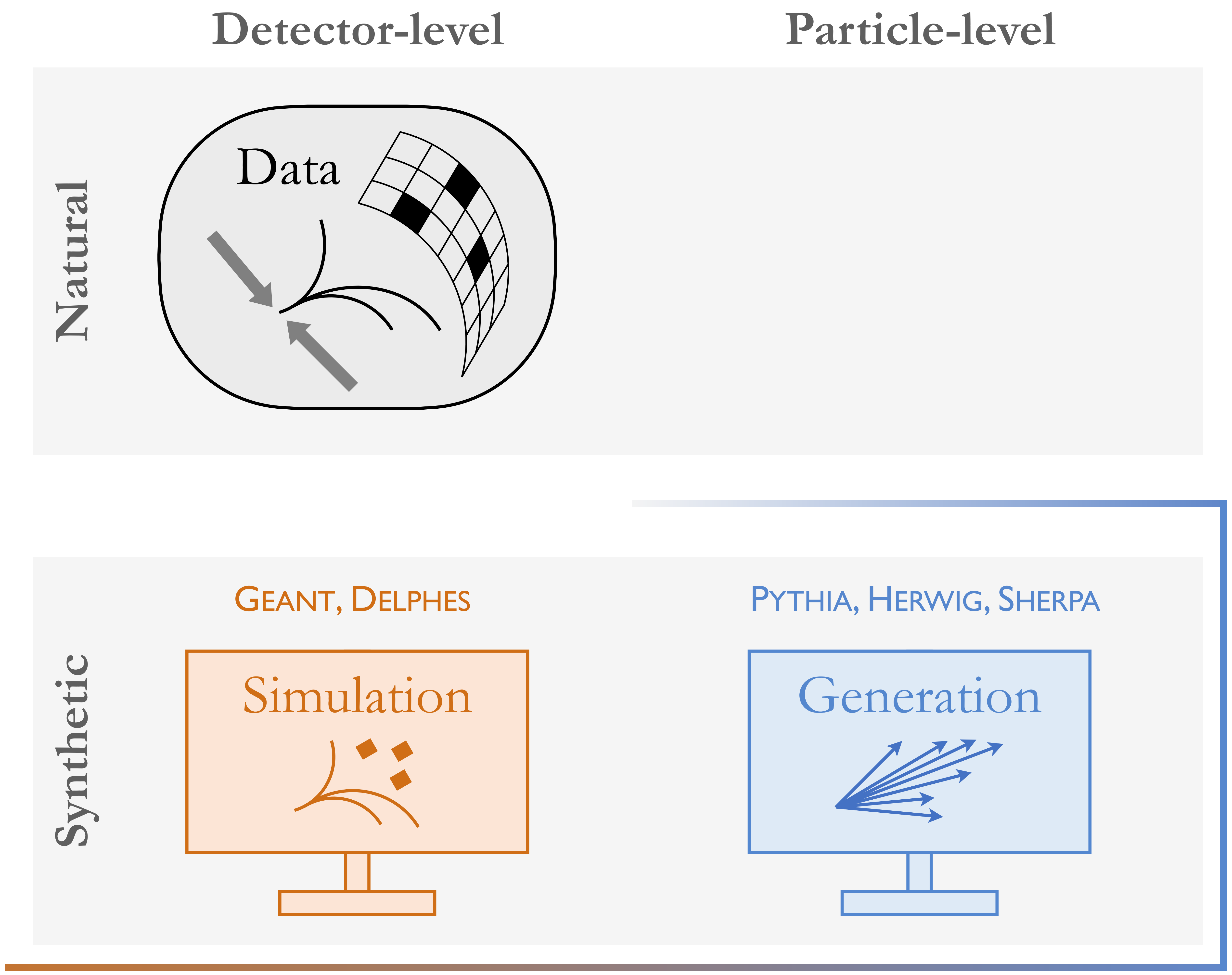
Correcting for Detector Effects – a.k.a. "Unfolding"

Measurements are affected by *detector effects* such as finite resolution, miscalibration, and limited acceptance



Correcting for Detector Effects – a.k.a. “Unfolding”

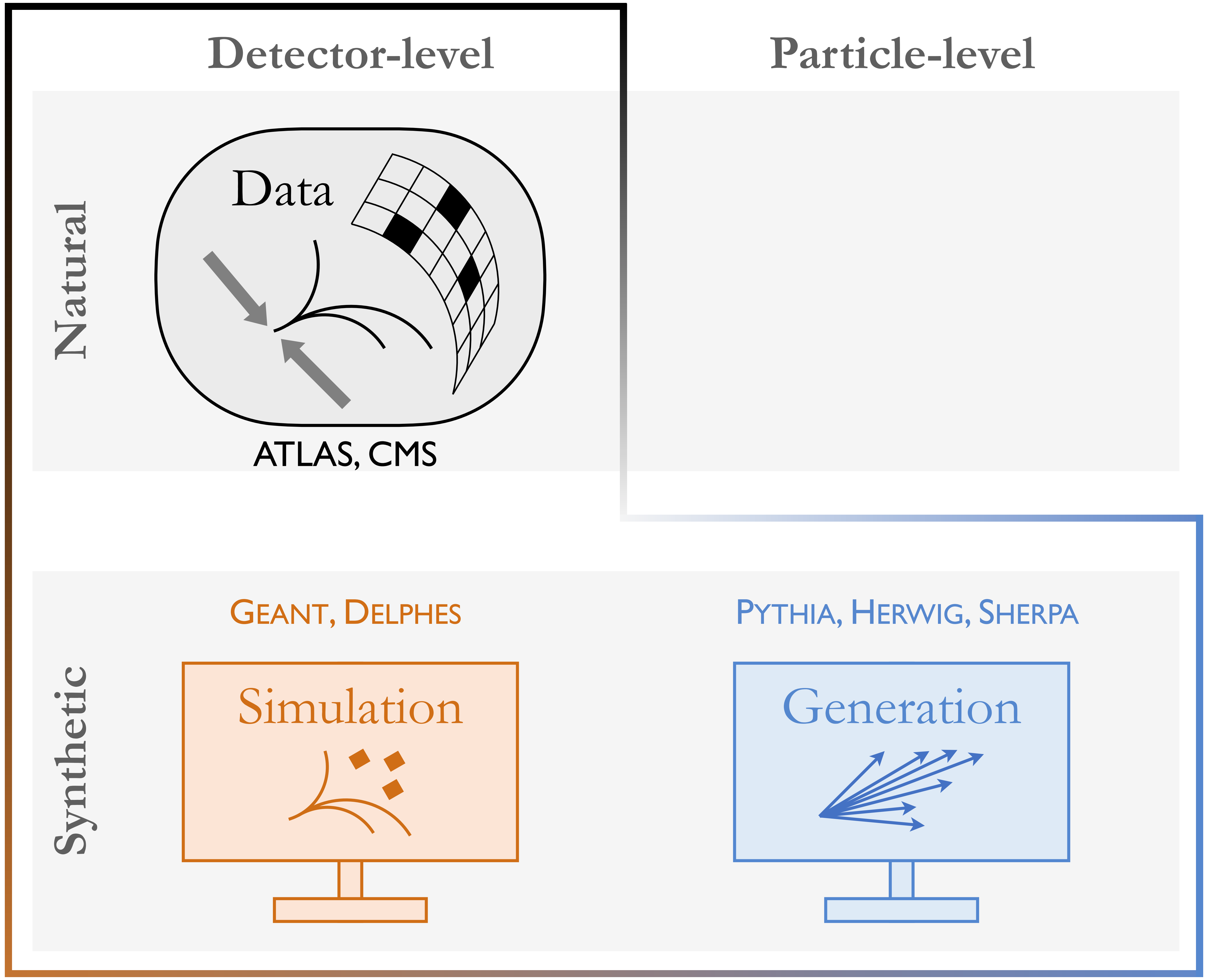
Measurements are affected by *detector effects* such as finite resolution, miscalibration, and limited acceptance



Learn detector response from *trustable* simulation

Correcting for Detector Effects – a.k.a. "Unfolding"

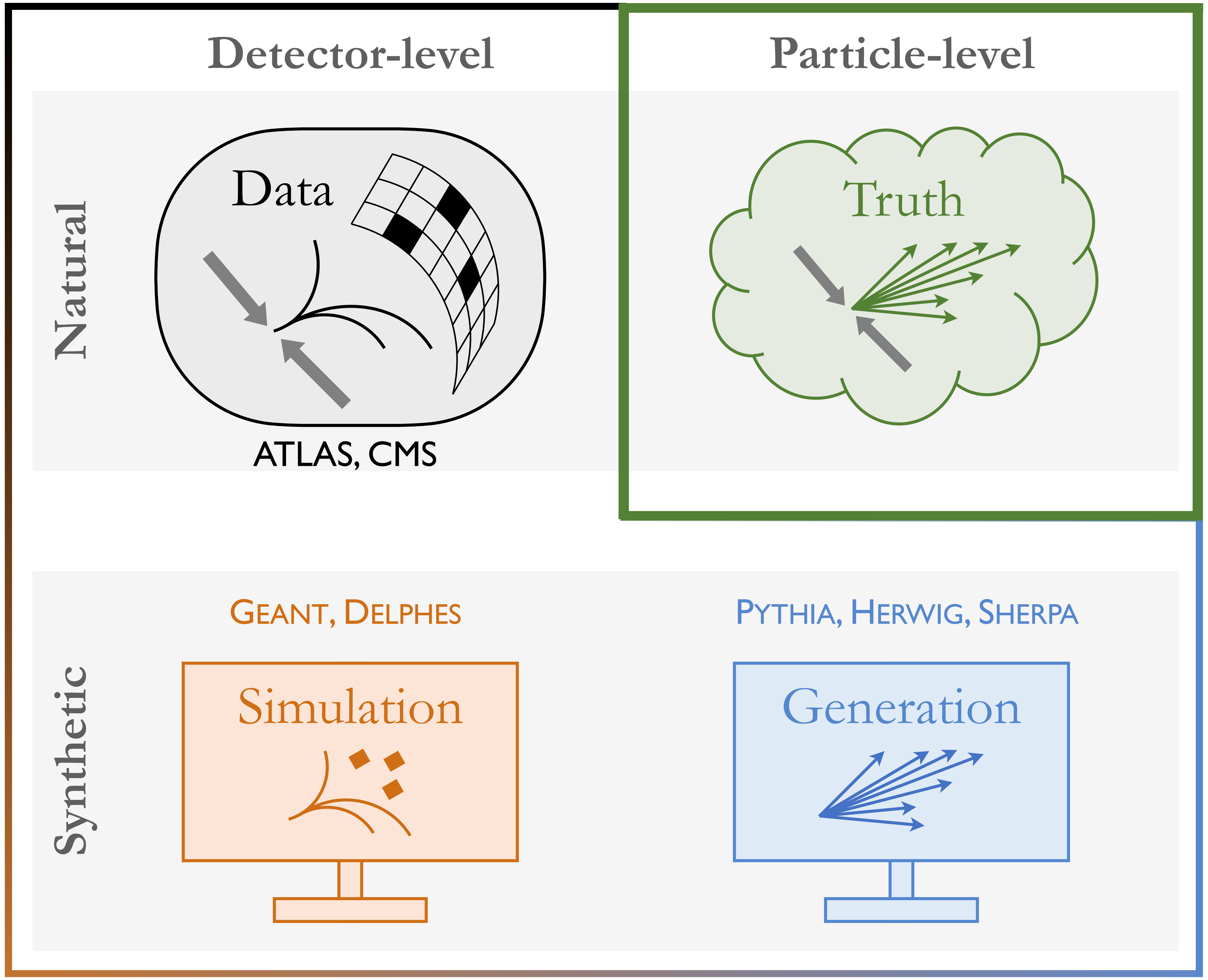
Measurements are affected by *detector effects* such as finite resolution, miscalibration, and limited acceptance



Learn detector response from *trustable* simulation

Correcting for Detector Effects – a.k.a. "Unfolding"

Measurements are affected by *detector effects* such as finite resolution, miscalibration, and limited acceptance



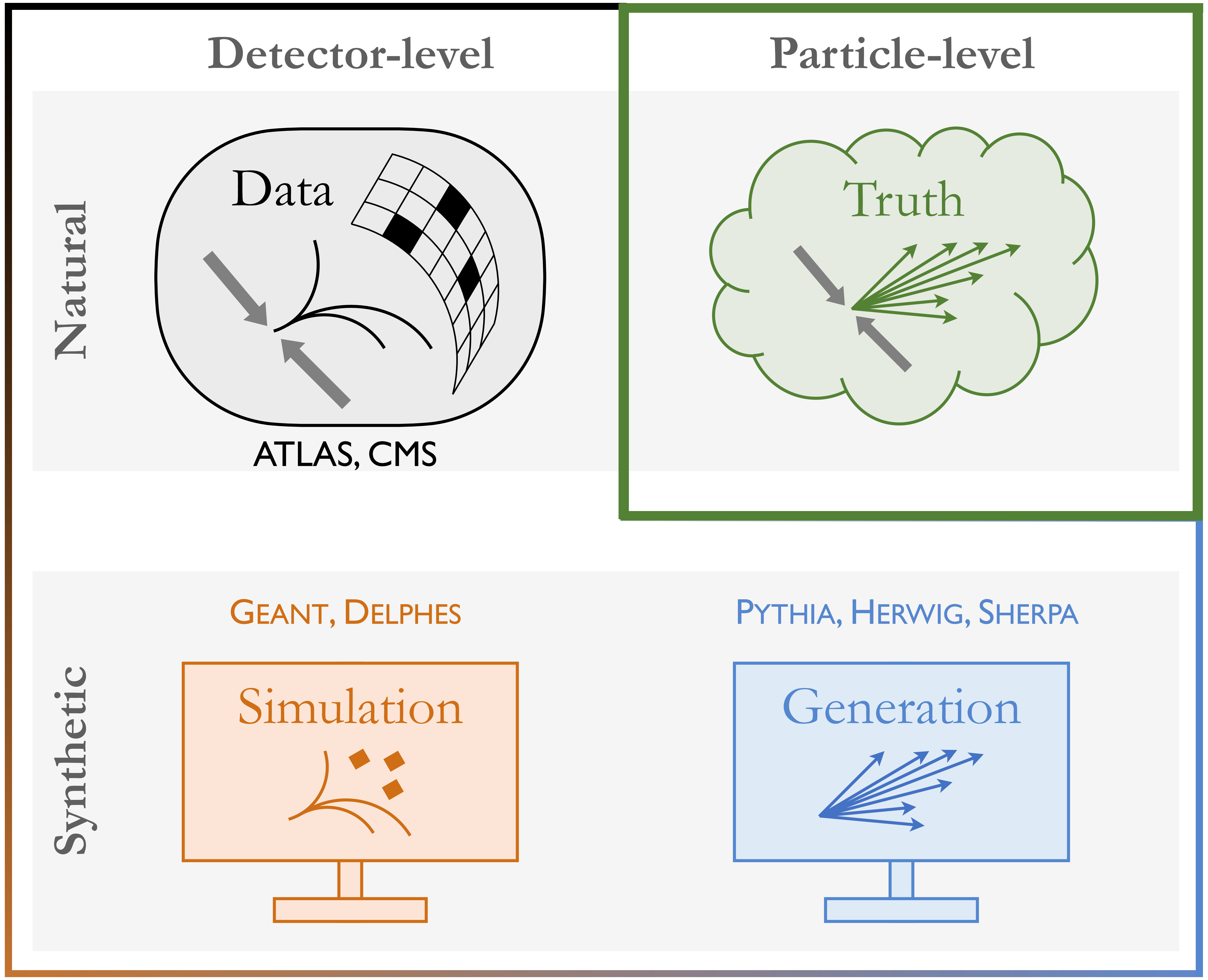
Learn detector response from *trustable* simulation

Truth-level measurements can be compared *across experiments* and to *theoretical calculations*

Goal of unfolding is to learn a particle-level model that reproduces the data

Correcting for Detector Effects – a.k.a. "Unfolding"

Measurements are affected by *detector effects* such as finite resolution, miscalibration, and limited acceptance



Learn detector response from *trustable simulation*

Truth-level measurements can be compared *across experiments* and to *theoretical calculations*

Goal of unfolding is to learn a particle-level model that reproduces the data

Unfolding Quantum Computer Readout Noise

[Nachman, Urbanen, de Jong, Bauer, 1910.01969]

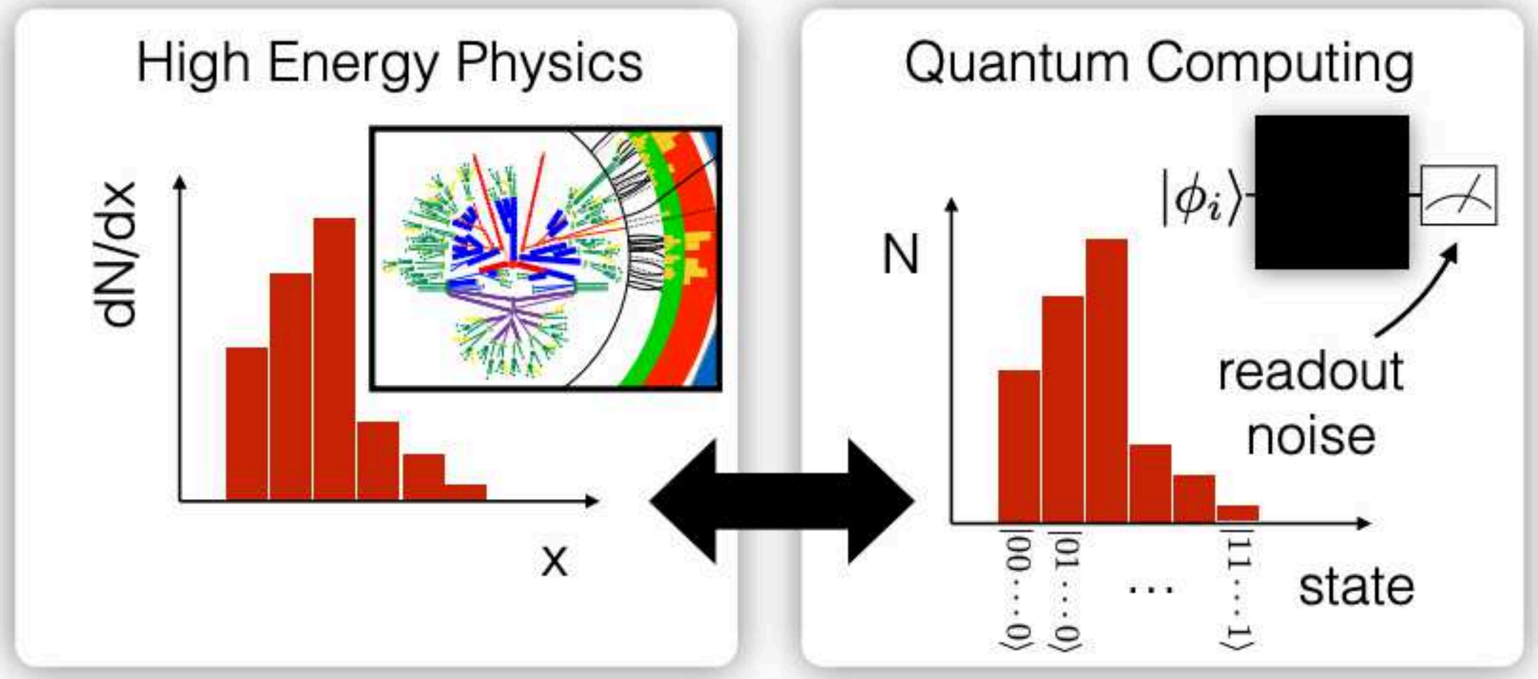


FIG. 1. A schematic diagram illustrating the connection between binned differential cross section measurements in high energy physics (left) and interpreting the output of repeated measurements from quantum computers (right).

Challenges with Traditional Unfolding

Previous methods explicitly rely on histograms

Binning fixed ahead of time, cannot be changed later
Performance of method sensitive to binning

Limited number of observables

Binning induces curse of dimensionality

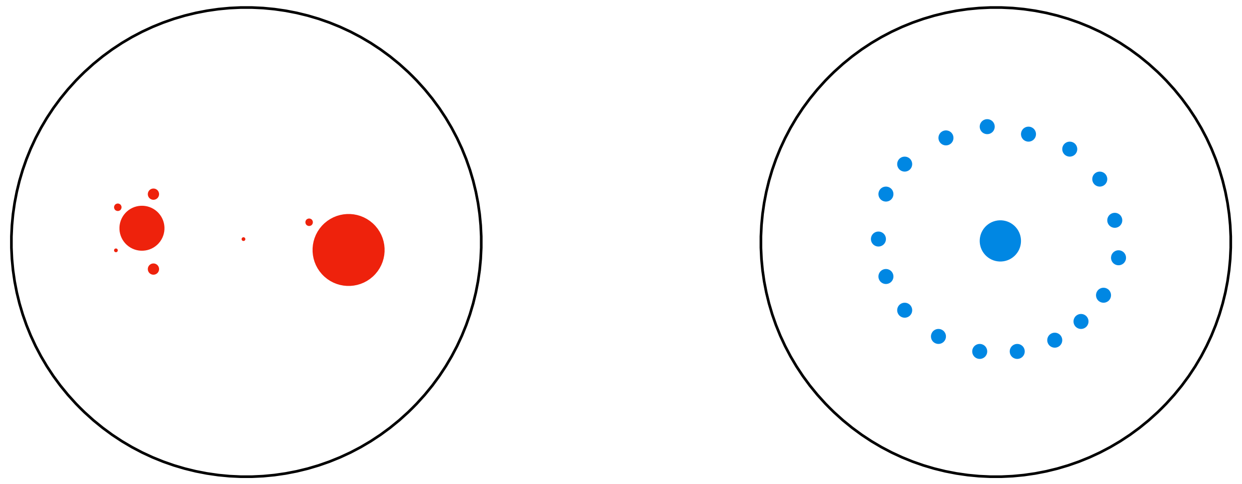
Response matrix depends on auxiliary features

Detector-level quantity may not capture full detector effect

Example – Two jets acquiring the same mass in different ways

Jet 1
Two hard prongs

Jet 2
Hard core, diffuse spray



Iterated Bayesian Unfolding (IBU)

also called Richardson-Lucy Deconvolution

Maximum likelihood, histogram-based unfolding method for a small number of observables

Choose observable(s) and binning at detector-level and particle-level

measured distribution: $m_i = \Pr(\text{measure } i)$

true distribution: $t_j^{(0)} = \Pr(\text{truth is } j)$

Calculate *response matrix* R_{ij} from generated/simulated pairs of events

$$R_{ij} = \Pr(\text{measure } i \mid \text{truth is } j)$$

Calculate new particle-level distribution using Bayes' theorem

$$t_j^{(n)} = \sum_i \Pr(\text{truth}_{n-1} \text{ is } j \mid \text{measure } i) \times \Pr(\text{measure } i)$$
$$= \sum_i \frac{R_{ij} t_j^{(n-1)}}{\sum_k R_{ik} t_k^{(n-1)}} \times m_i$$

Iterate procedure to remove dependence on prior

[Richardson, JOSA 1972; Lucy, AJ 1974; D'Agostini, NIMPA 1995]

Likelihood Reweighting via Classification

Likelihood ratio is optimal binary classifier by Neyman-Pearson lemma

$$L[(w, X), (w', X')](x) = \frac{p_{(w,X)}(x)}{p_{(w',X')}(x)}$$

L – likelihood ratio
w – weights
X – phase space
x – element of X
p – probability density

Likelihood Reweighting via Classification

Likelihood ratio is optimal binary classifier by Neyman-Pearson lemma

$$L[(w, X), (w', X')](x) = \frac{p_{(w,X)}(x)}{p_{(w',X')}(x)}$$

L – likelihood ratio
w – weights
X – phase space
x – element of X
p – probability density

Model output of a well-trained classifier accesses likelihood ratio

$$\text{Model}[(w, X), (w', X')](x) \simeq \frac{L[(w, X), (w', X')](x)}{1 + L[(w, X), (w', X')](x)}$$

Assuming softmax output, categorical cross-entropy loss

[Cranmer, Pavez, Louppe, 1506.02169; Andreassen, Nachman, PRD 2020]

Likelihood Reweighting via Classification

Likelihood ratio is optimal binary classifier by Neyman-Pearson lemma

$$L[(w,X),(w',X')](x) = \frac{p_{(w,X)}(x)}{p_{(w',X')}(x)}$$

L – likelihood ratio
w – weights
X – phase space
x – element of X
p – probability density

Model output of a well-trained classifier accesses likelihood ratio

$$\text{Model}[(w,X),(w',X')](x) \simeq \frac{L[(w,X),(w',X')](x)}{1+L[(w,X),(w',X')](x)}$$

Assuming softmax output, categorical cross-entropy loss

[Cranmer, Pavez, Louppe, 1506.02169; Andreassen, Nachman, PRD 2020]

OmniFold repeatedly reweights one weighted sample (A) to another (B)

$$w_{A'}(x) = w_A(x) \times \frac{\text{Model}[(w_B,B),(w_A,A)](x)}{1-\text{Model}[(w_B,B),(w_A,A)](x)}$$

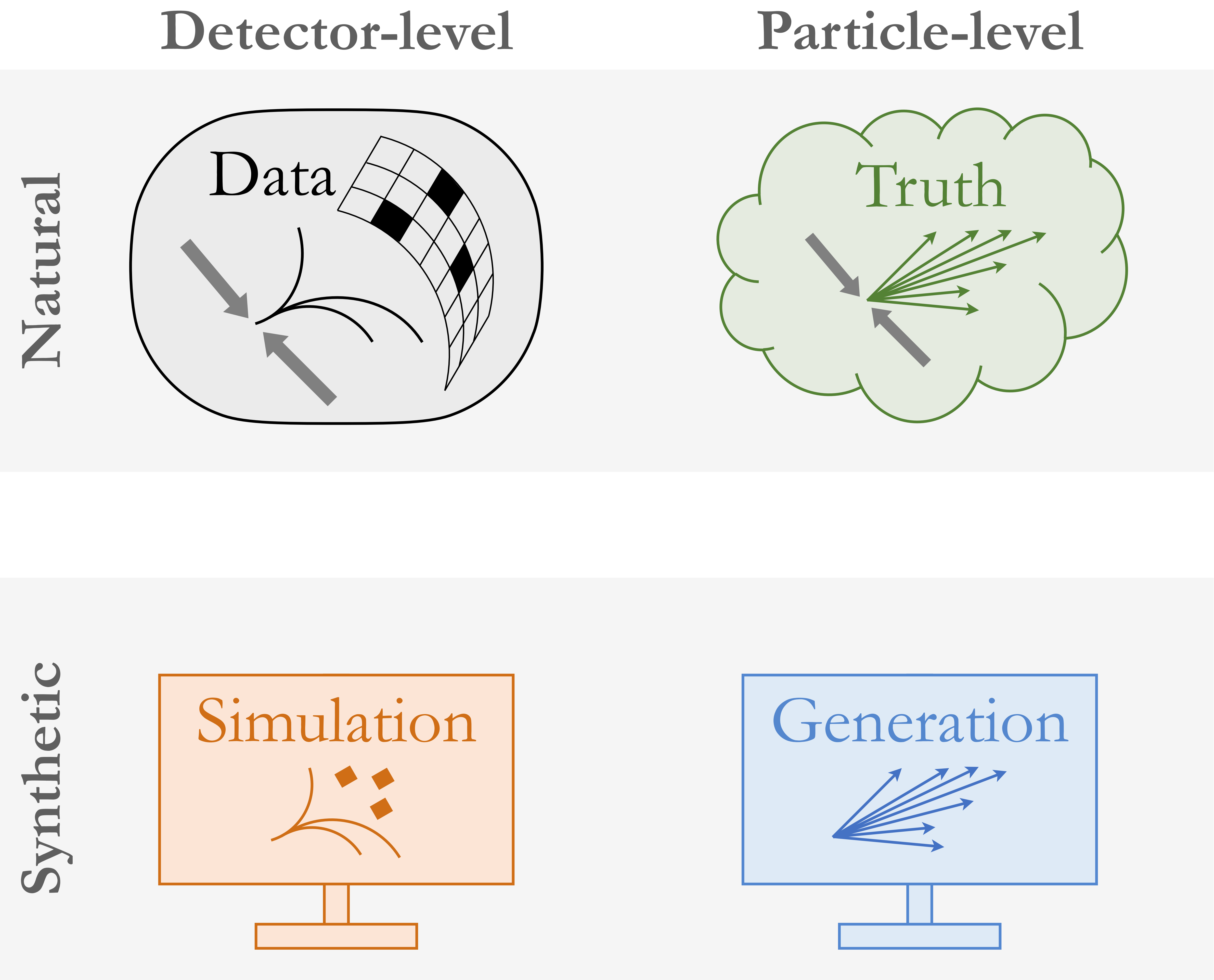
A' is statistically indistinguishable from B

Likelihood reweighting requires effective classification of events

Simultaneously Unfolding All Observables – OmniFold

OmniFold weights particle-level Gen to be consistent with Data once passed through the detector

[Andreassen, PTK, Metodiev, Nachman, Thaler, PRL 2020]



Simultaneously Unfolding All Observables – OmniFold

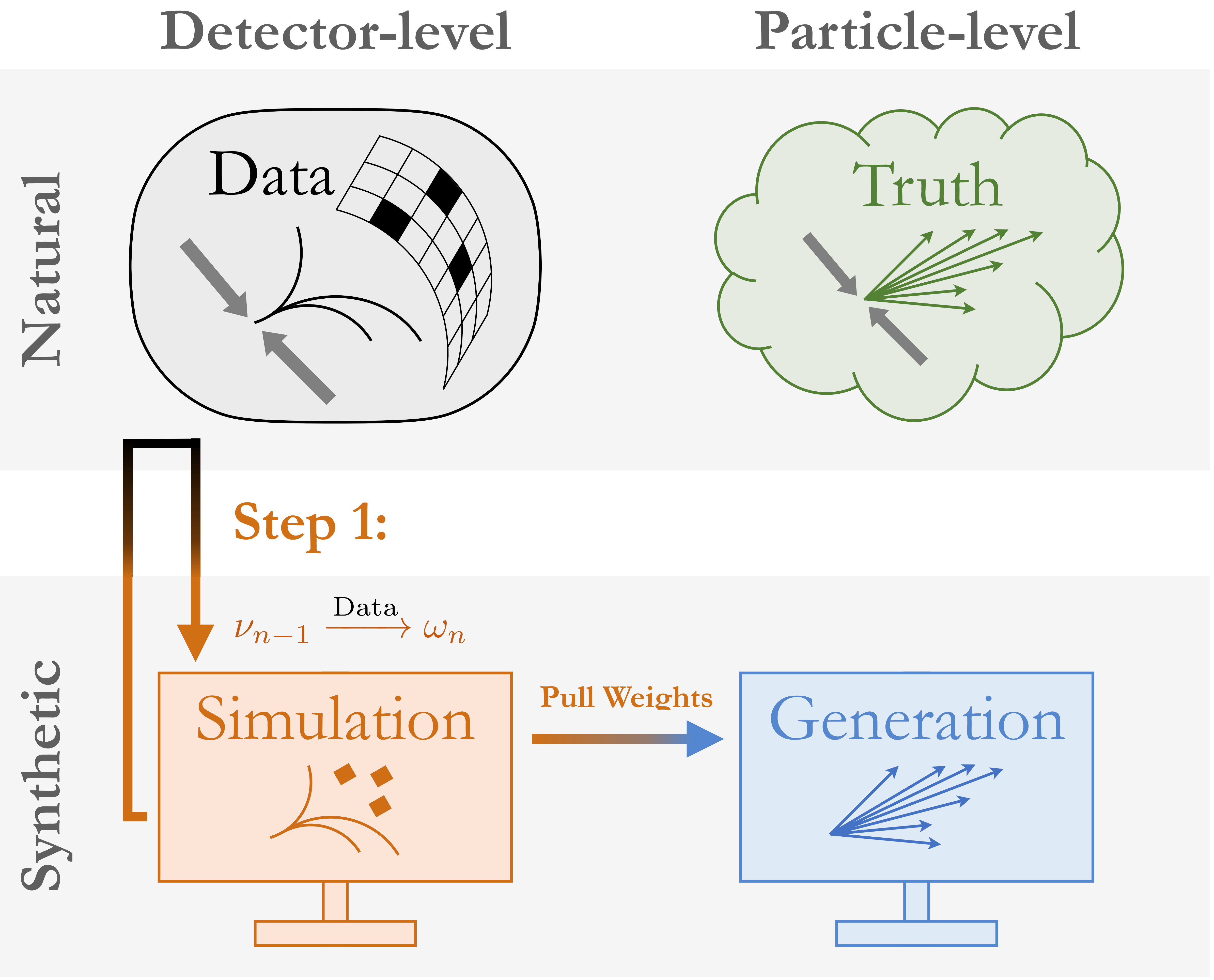
[Andreassen, PTK, Metodiev, Nachman, Thaler, PRL 2020]

OmniFold weights particle-level Gen to be consistent with Data once passed through the detector

Step 1

Reweights Sim_{n-1} to data, pulls weights back to particle-level Gen_{n-1}

Incorporates the response matrix



Simultaneously Unfolding All Observables – OmniFold

[Andreassen, PTK, Metodiev, Nachman, Thaler, PRL 2020]

OmniFold weights particle-level Gen to be consistent with Data once passed through the detector

Step 1

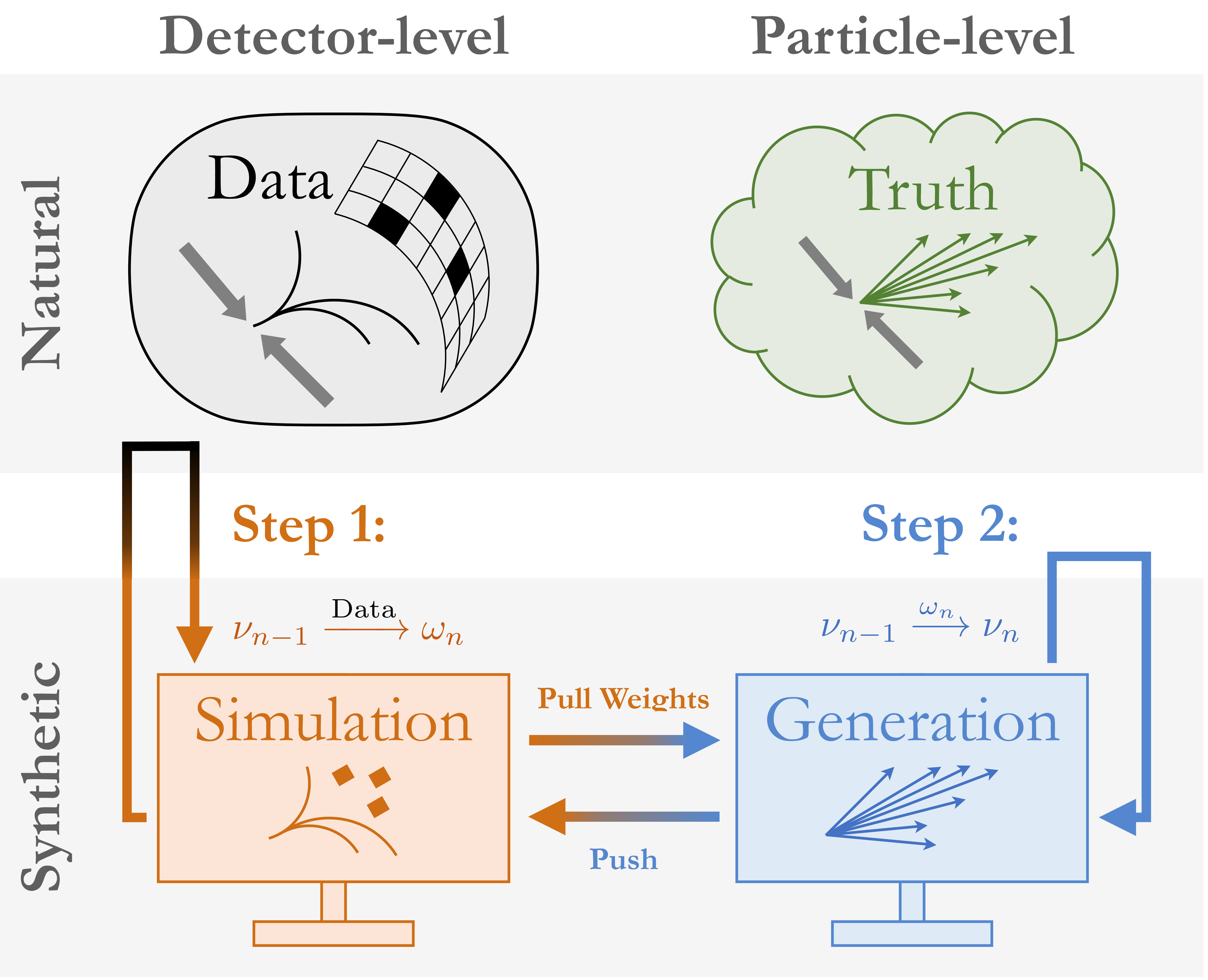
Reweights Sim_{n-1} to data, pulls weights back to particle-level Gen_{n-1}

Incorporates the response matrix

Step 2

Reweights Gen_{n-1} to (step 1)-weighted gen_{n-1}, pushes weights to detector-level Sim_n

Constructs valid particle-level function by averaging gen-level weights



Simultaneously Unfolding All Observables – OmniFold

[Andreassen, PTK, Metodiev, Nachman, Thaler, PRL 2020]

OmniFold weights particle-level Gen to be consistent with Data once passed through the detector

Step 1

Reweights Sim_{n-1} to data, pulls weights back to particle-level Gen_{n-1}

Incorporates the response matrix

Step 2

Reweights Gen_{n-1} to (step 1)-weighted gen_{n-1}, pushes weights to detector-level Sim_n

Constructs valid particle-level function by averaging gen-level weights

OmniFold

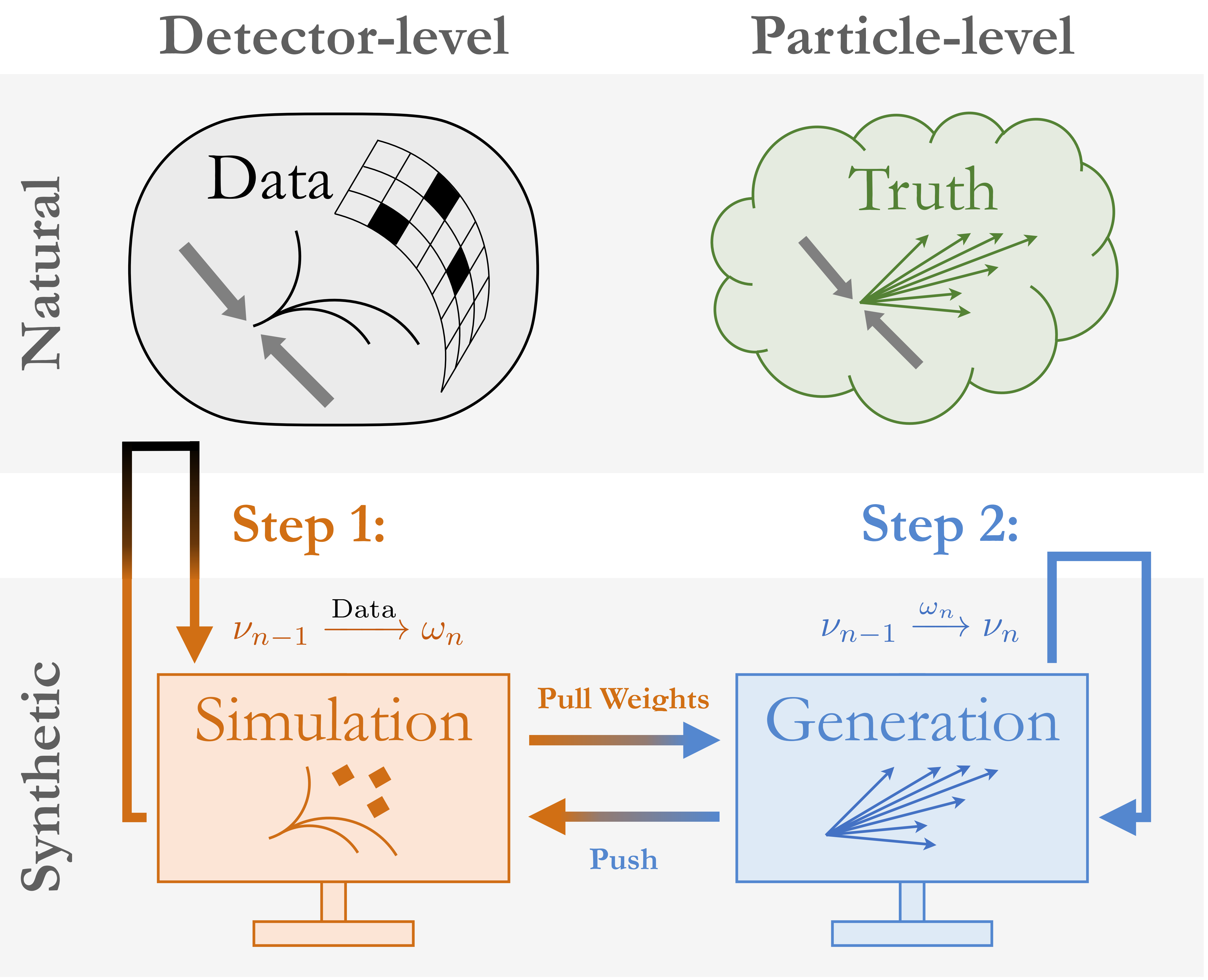
Step 1 – $\omega_n(m) = \nu_{n-1}^{\text{push}} \times L[(1, \text{Data}), (\nu_{n-1}^{\text{push}}, \text{Sim})](m)$

Step 2 – $\nu_n(t) = \nu_{n-1}(t) \times L[(\omega_n^{\text{pull}}, \text{Gen}), (\nu_{n-1}, \text{Gen})](t)$

Unfold *any** observable $p_{\text{Gen}}(t)$ using universal weights $\nu_n(t)$

$$p_{\text{unfolded}}^{(n)}(t) = \nu_n(t) \times p_{\text{Gen}}(t)$$

*Observables should be chosen responsibly



Simultaneously Unfolding All Observables – OmniFold

OmniFold weights particle-level Gen to be consistent

[Andreassen, PTK, Metodiev, Nachman, Thaler, PRL 2020]

Rew

Re

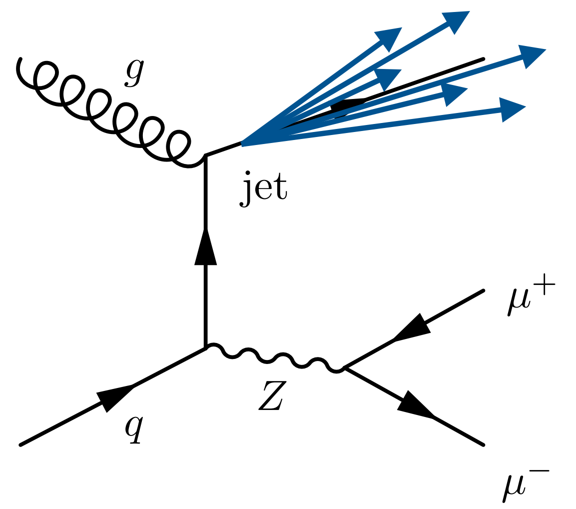
Co

The Mountain sat upon the Plain
In his tremendous Chair –
His observation **OmniFold**,
His inquest, everywhere –

The Seasons played around his knees
Like Children round a sire –
Grandfather of the Days is He
Of Dawn, the Ancestor –

Emily Dickinson, #975

Testing OmniFold – Z + Jet Case Study



[Andreassen, PTK, Metodiev, Nachman, Thaler, PRL 2020]

$Z(\to \mu^+\mu^-)$ + Jet Events
- "Data" – HERWIG 7.1.5
- MC – PYTHIA 8.243, tune 26
- 1.6 million events each after cuts

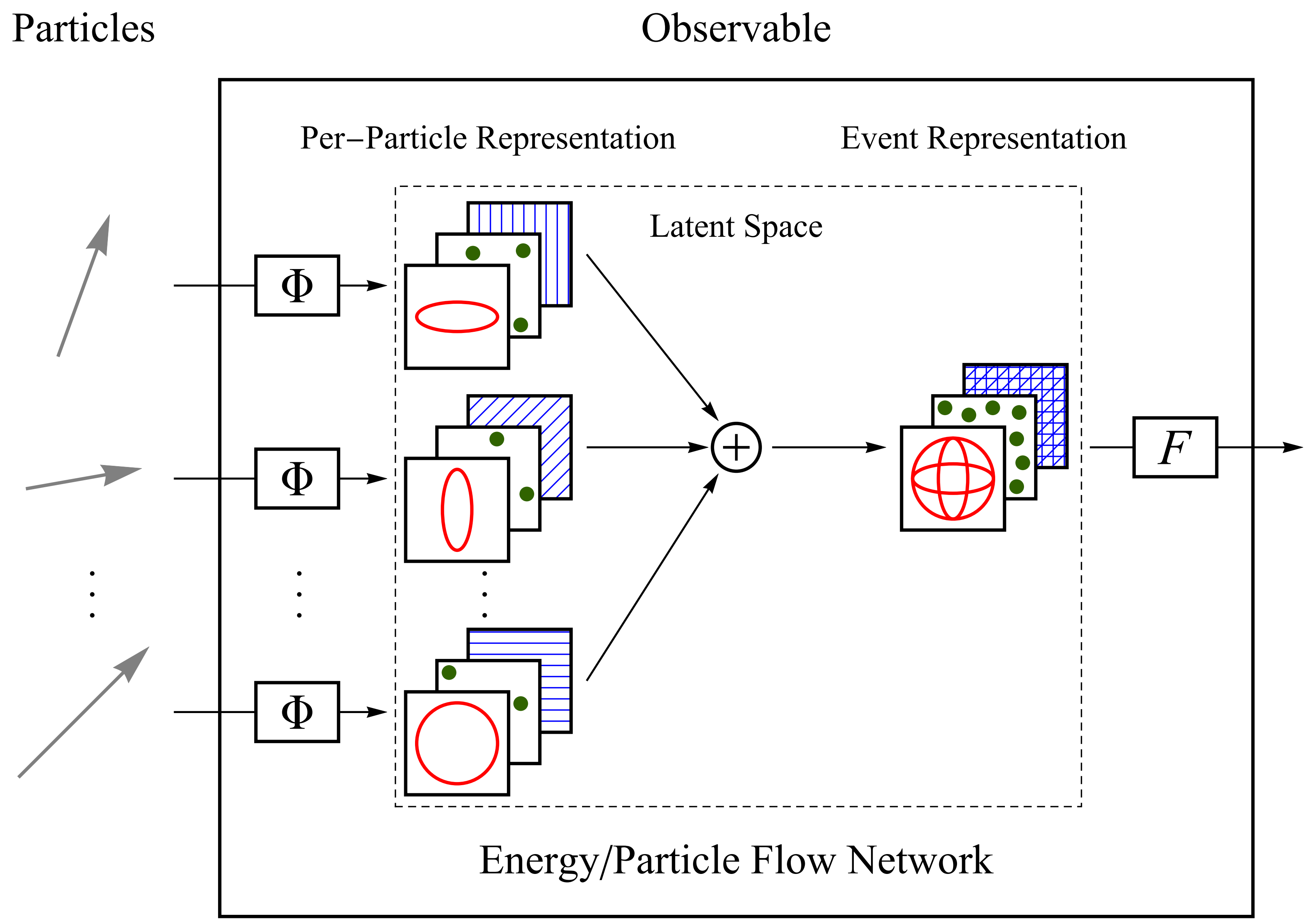
Detector Simulation
- CMS-like detector – DELPHES 3.4.2

Jets
- Anti-$k_T, R = 0.4$ – FASTJET 3.3.2
- $p_T^Z > 200$ GeV, assume excellent muon detector resolution

Datasets publicly available
- *With two additional Pythia tunes*
- *Accessible via EnergyFlow*

OmniFold Binder Demo

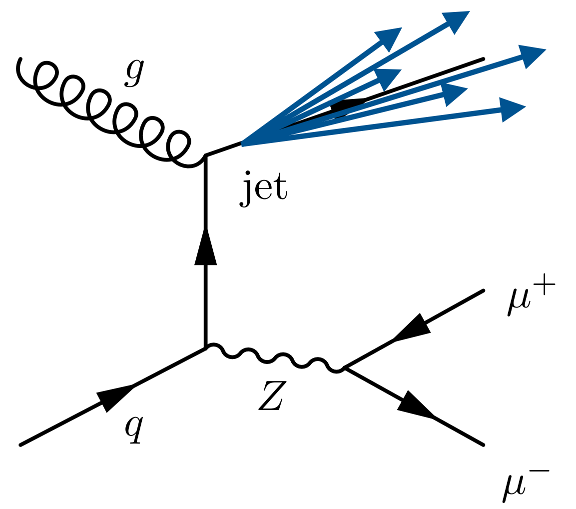
[PTK, Metodiev, Thaler, JHEP 2019]



Particle Flow Network (PFN) architecture processes full radiation pattern of the event

- PFN-Ex: $(p_T, y, \phi, \text{PID})$ input features
- $\Phi:\ (100, 100, 256)$ dense layers
- $F:\ (100,100,100)$ dense layers
- ReLU activations, softmax output
- Categorical cross-entropy loss
- 20% validation sample
- 10 epoch patience

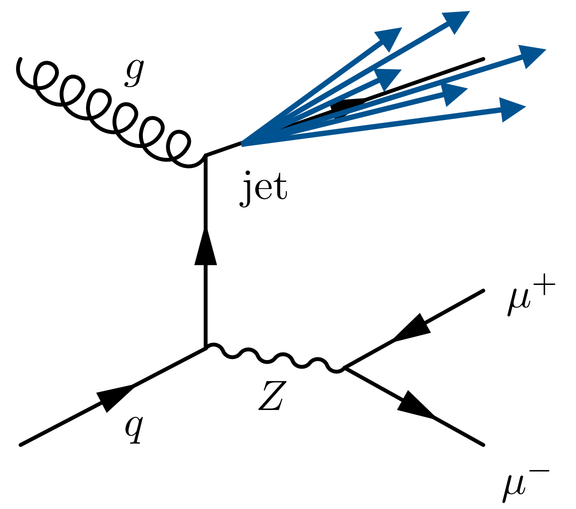
OmniFolding Jet Substructure Observables



Single OmniFold instantiation vs. separate instantiations of IBU

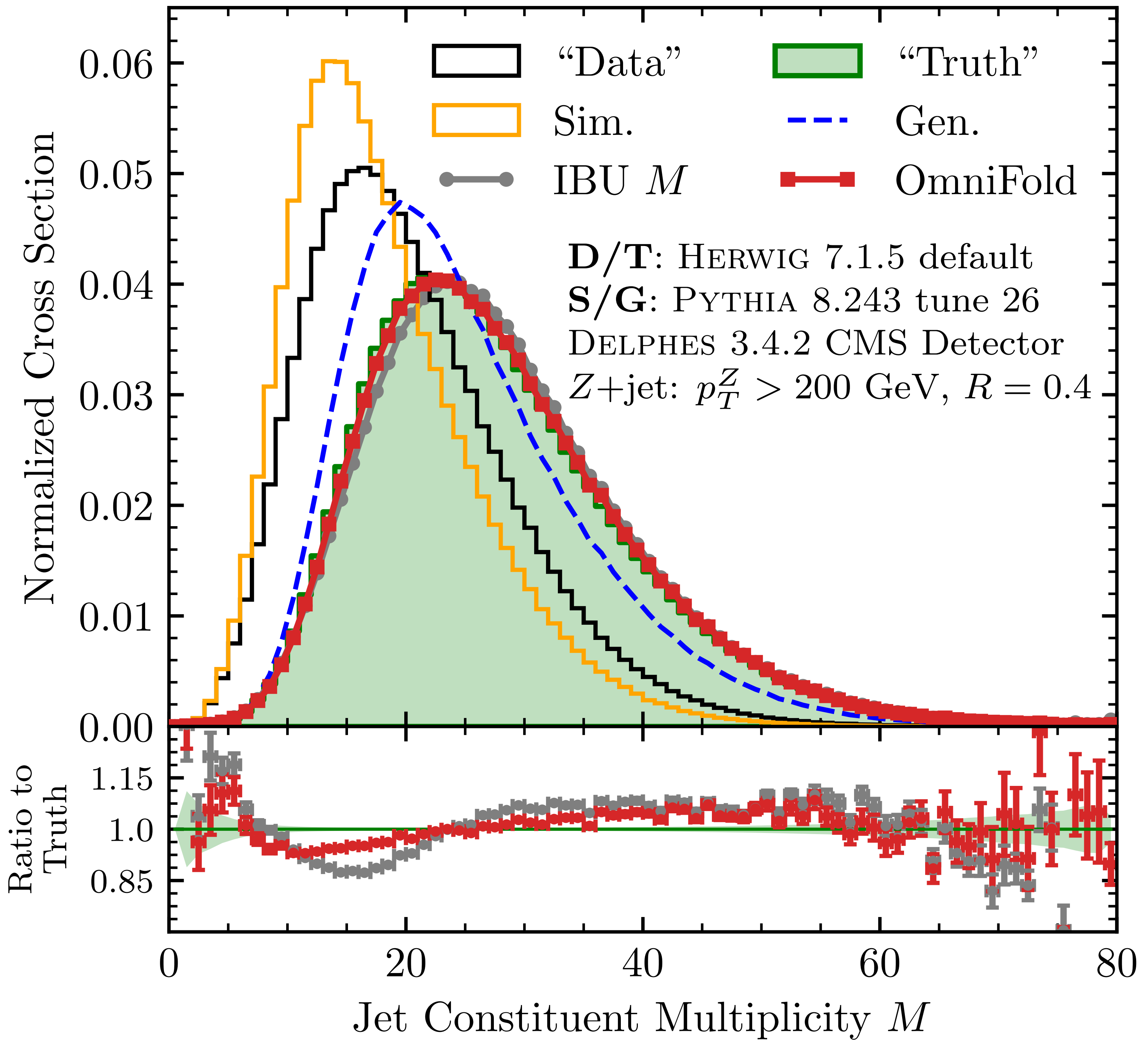
Successful unfolding means IBU/OmniFold should approach Truth

OmniFolding Jet Substructure Observables



Single OmniFold instantiation vs. separate instantiations of IBU

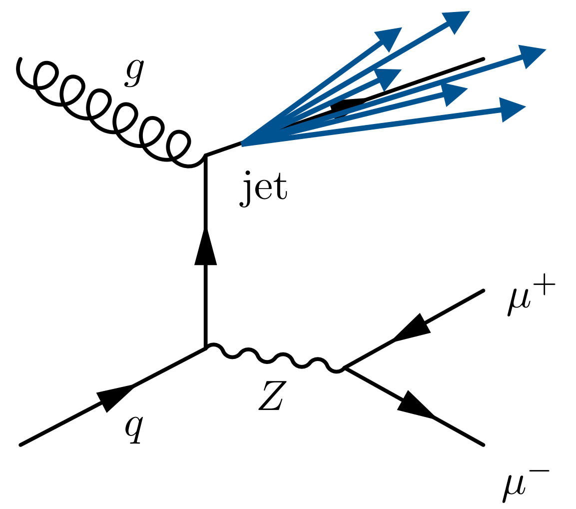
Successful unfolding means IBU/OmniFold should approach Truth



Number of particles in the jet

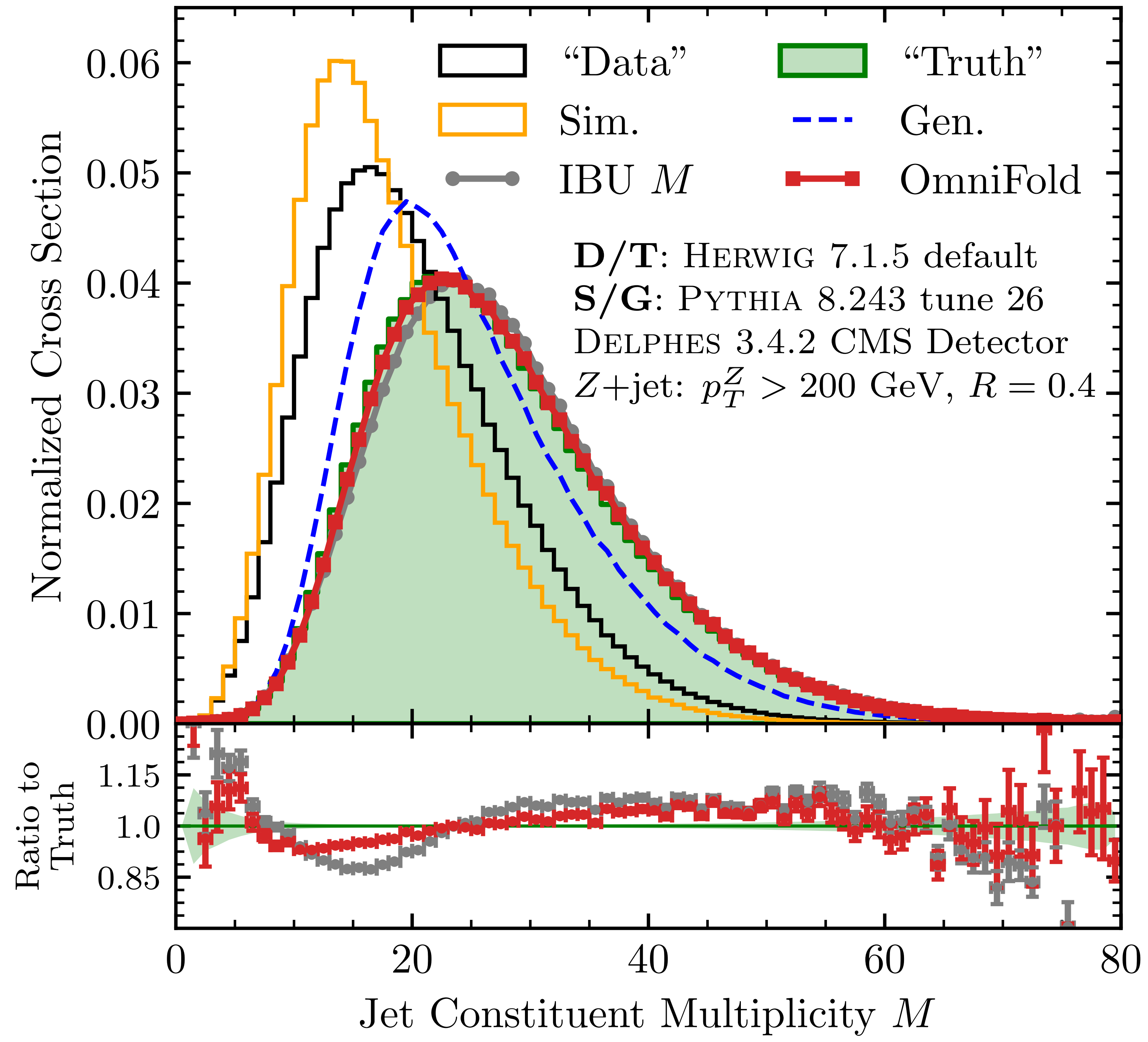
OmniFold equals or outperforms IBU

OmniFolding Jet Substructure Observables

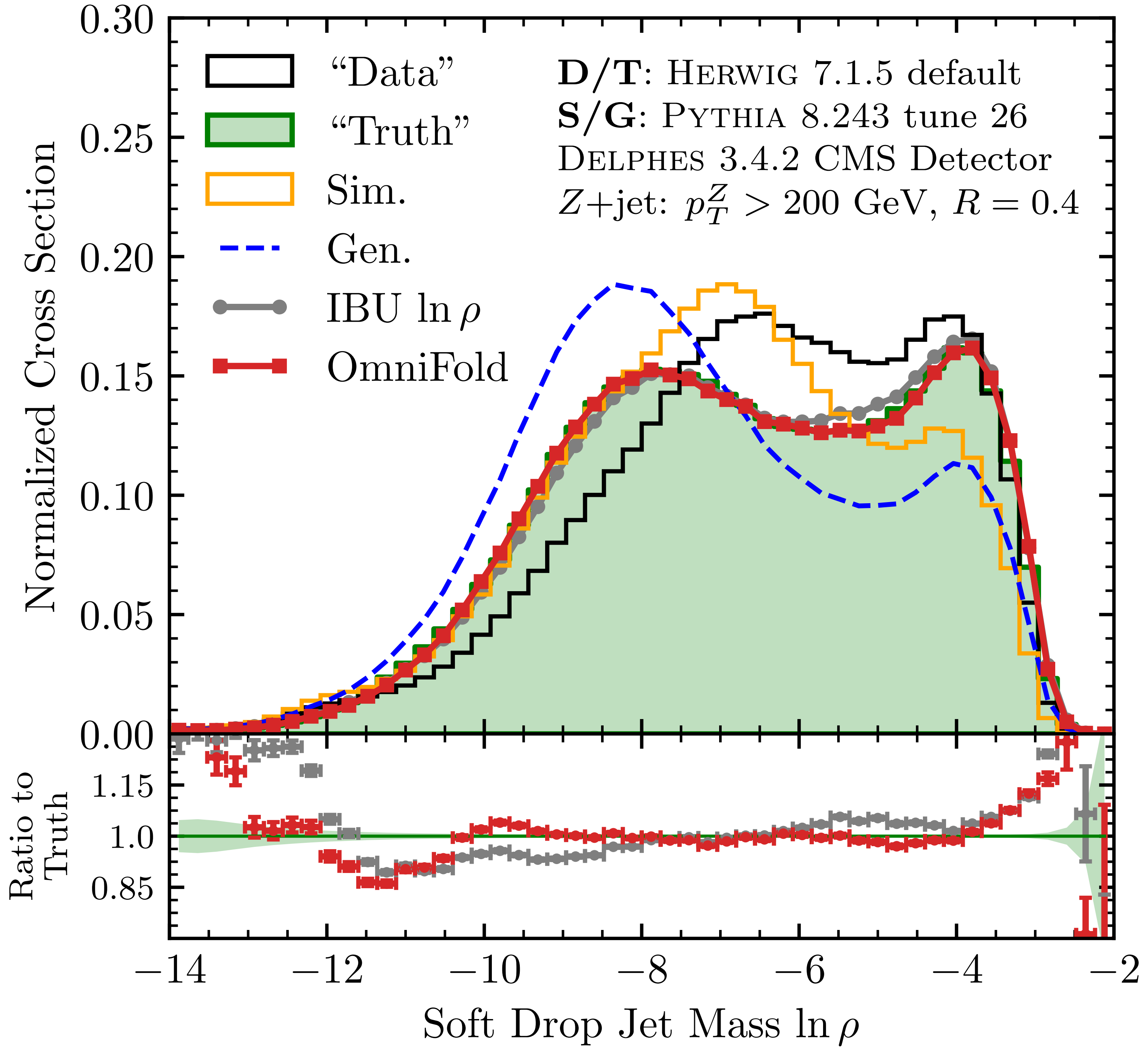


Single OmniFold instantiation vs. separate instantiations of IBU

Successful unfolding means IBU/OmniFold should approach Truth



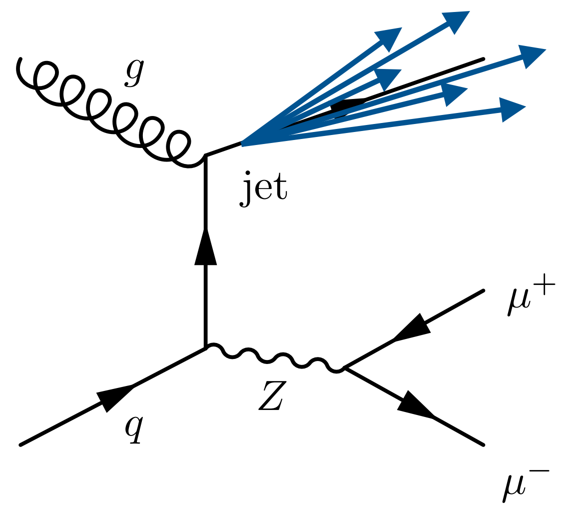
Number of particles in the jet



De-noised invariant mass of jet

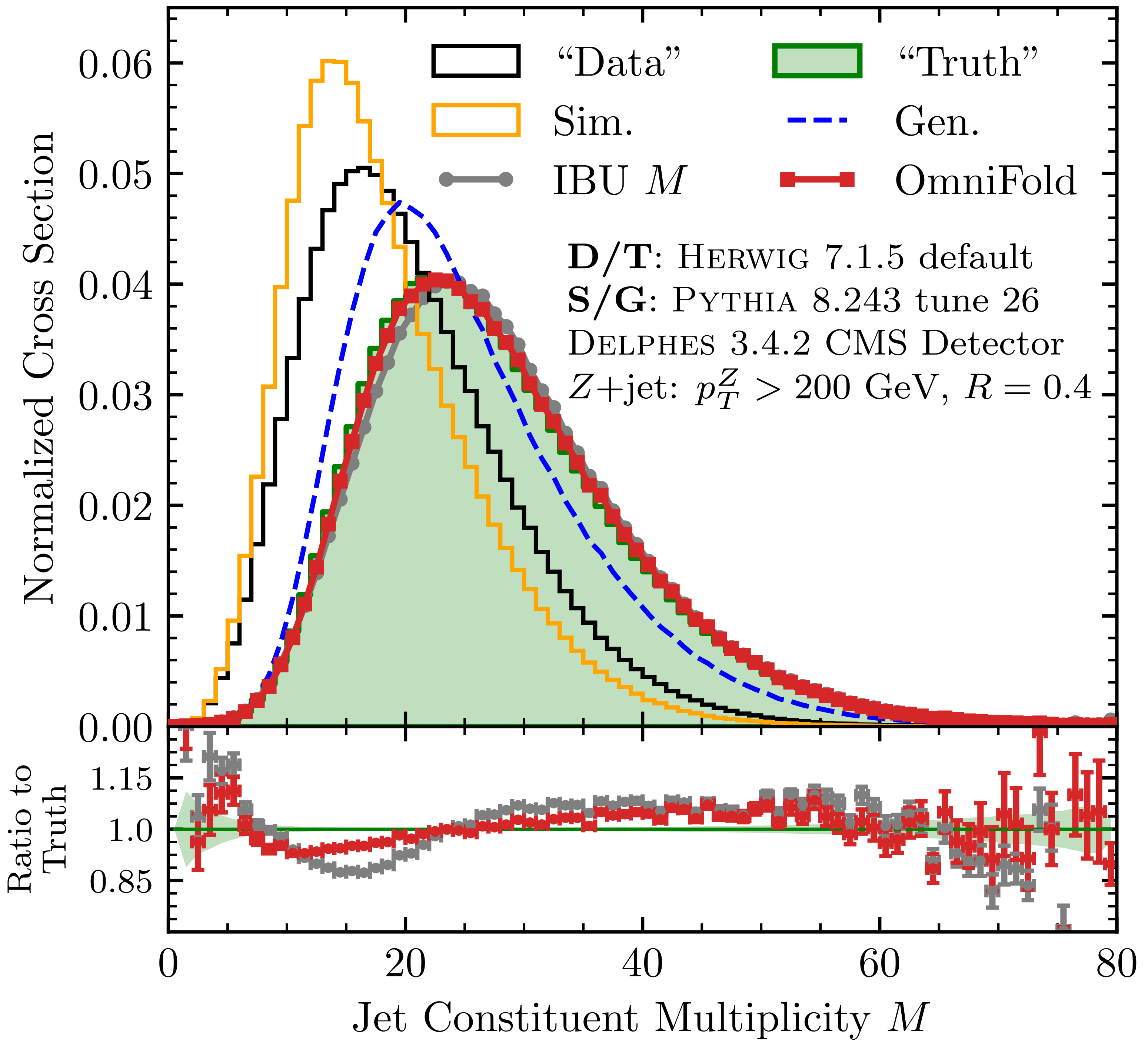
OmniFold equals or outperforms IBU

OmniFolding Jet Substructure Observables

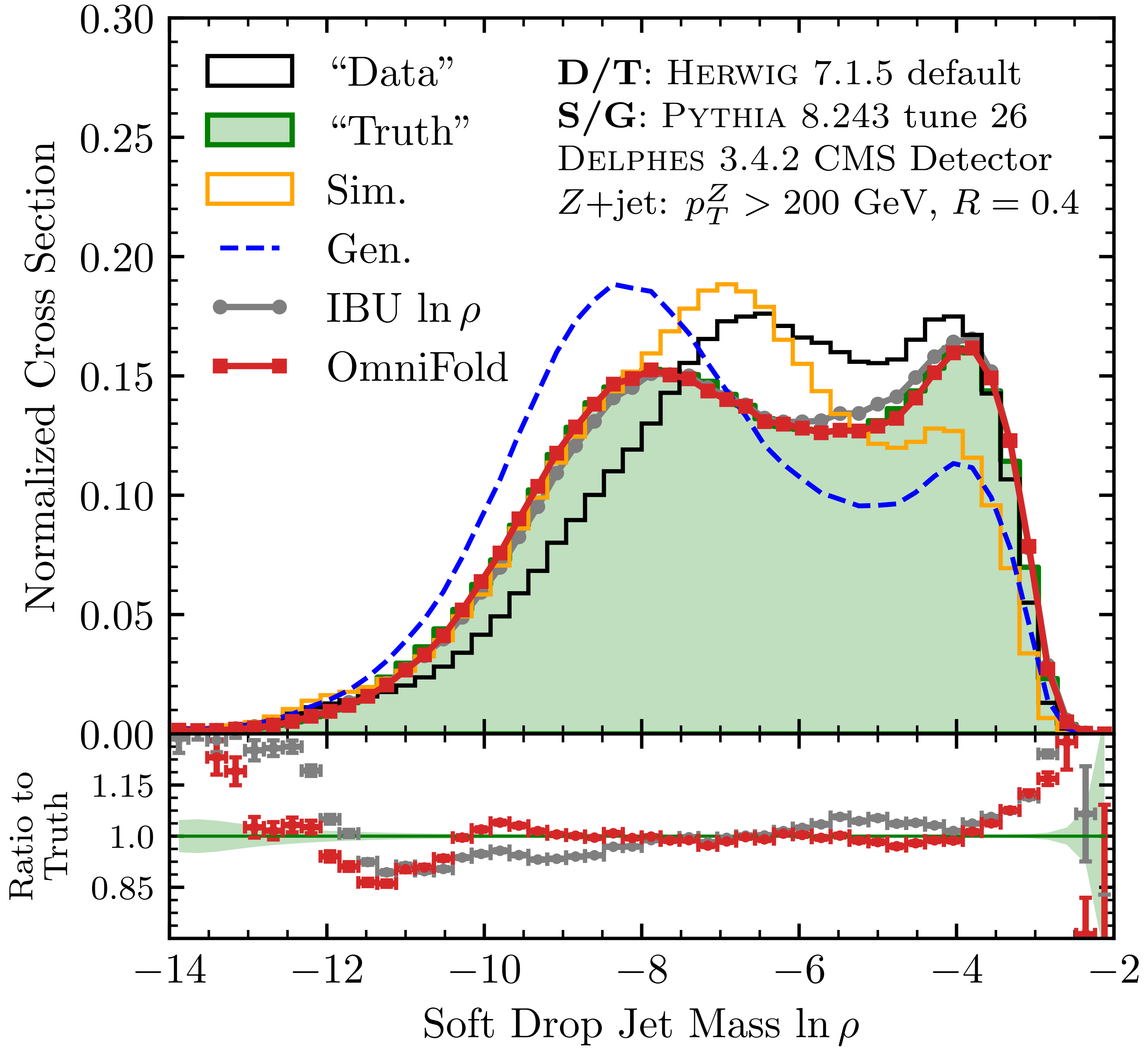


Single OmniFold instantiation vs. separate instantiations of IBU

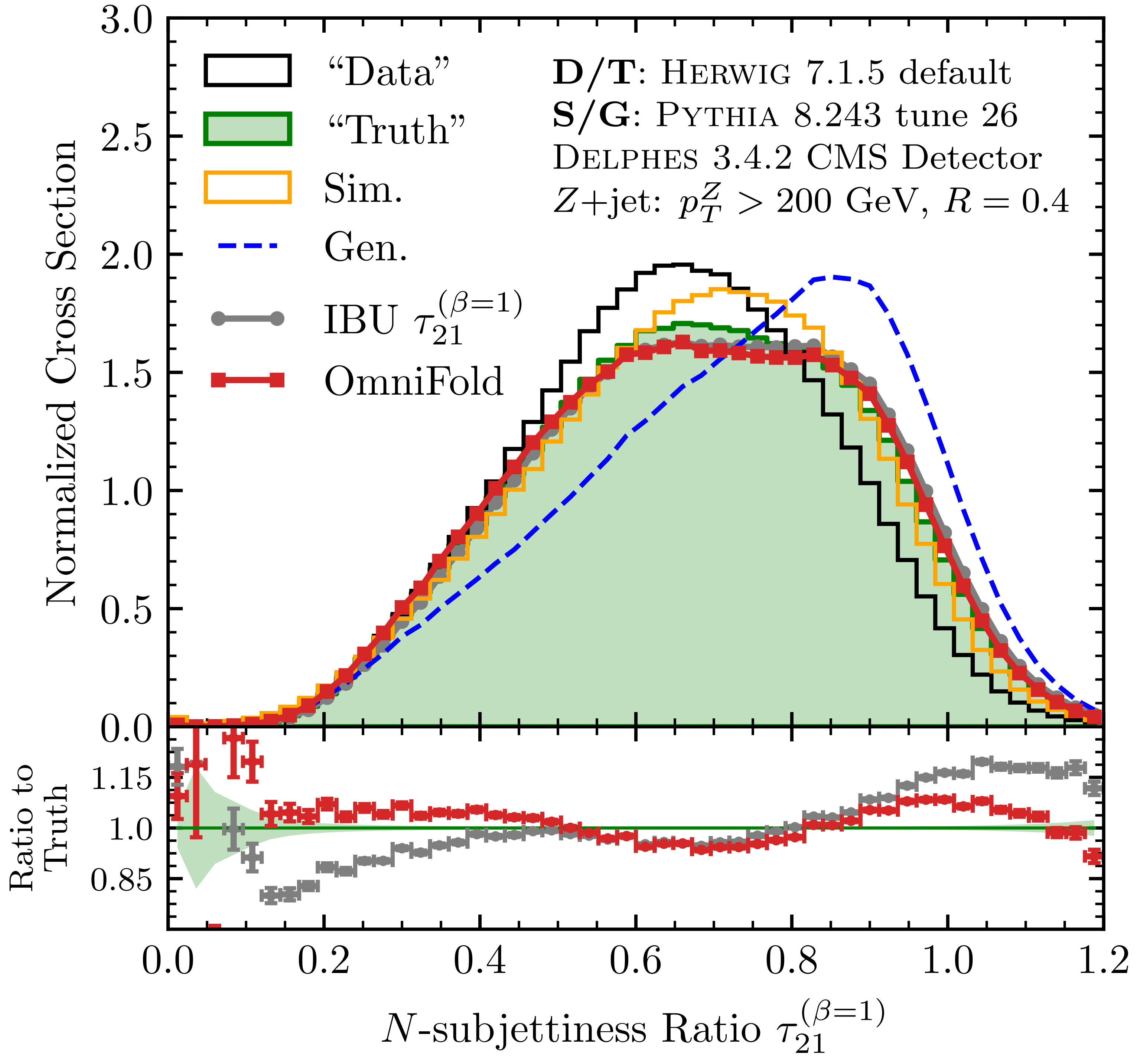
Successful unfolding means IBU/OmniFold should approach Truth



Number of particles in the jet



De-noised invariant mass of jet

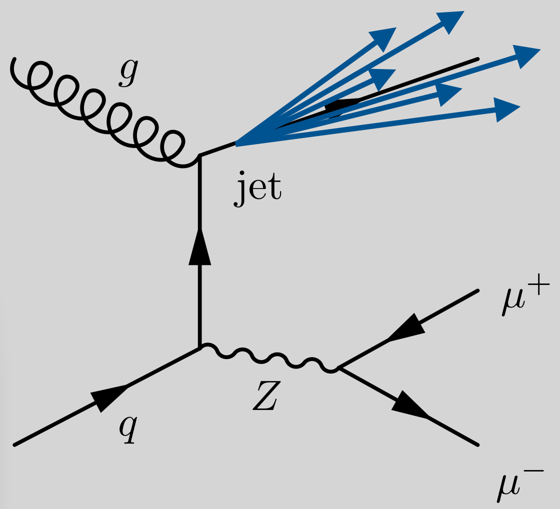


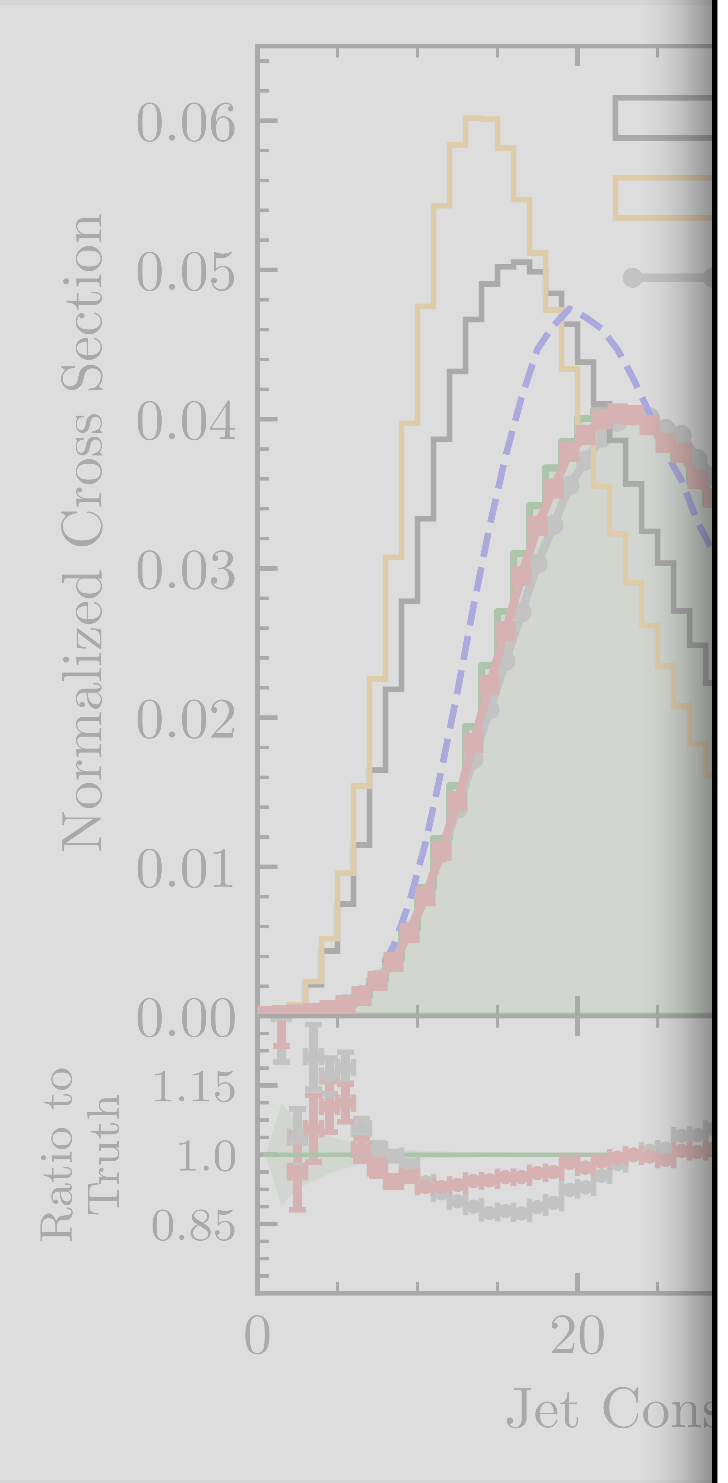
Ratio of "two-pronginess" to "one-pronginess"

OmniFold equals or outperforms IBU

(Additional unfolded distributions in backup)

OmniFolding Jet Substructure Observables





Number

Measurement of lepton-jet correlations in high Q^2 neutral-current DIS with the H1 detector at HERA

The H1 Collaboration

H1prelim-21-031
April 4, 2021

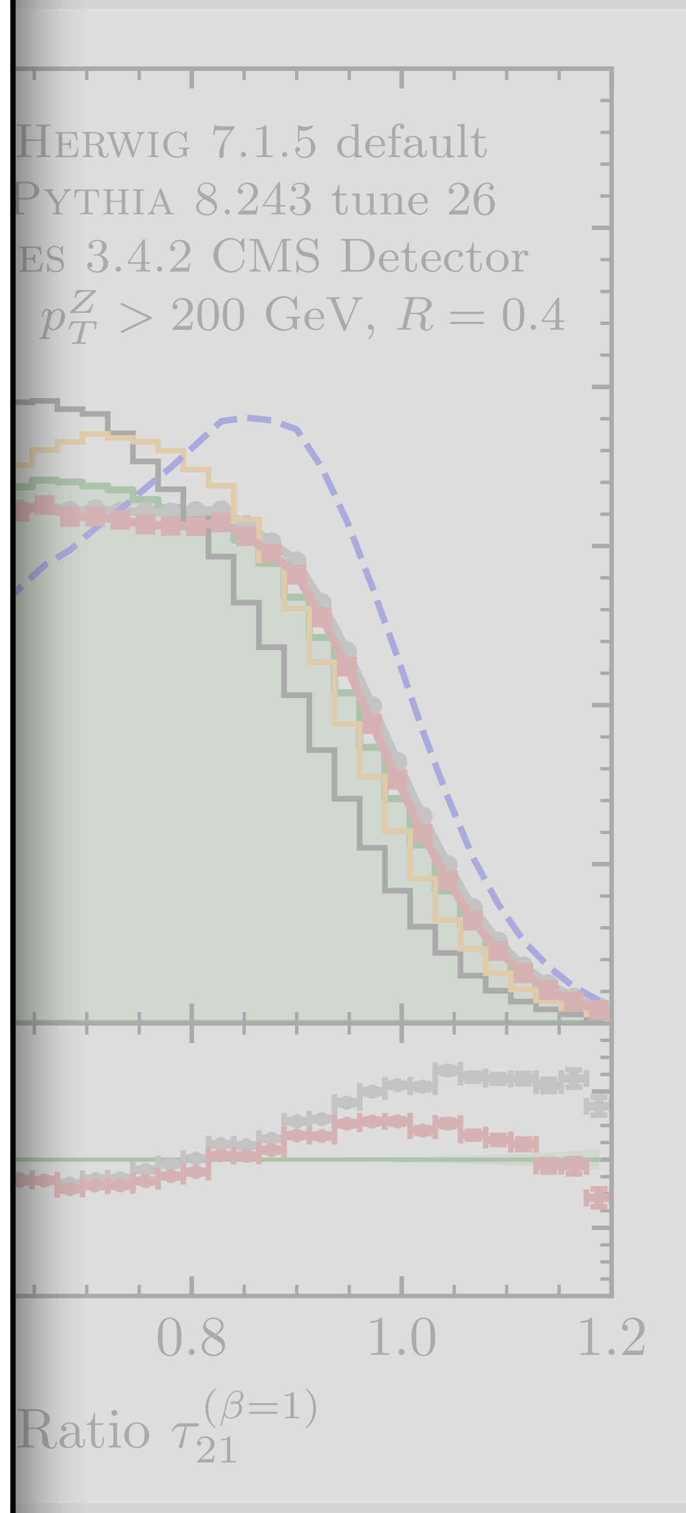
Abstract

A measurement of jet production in high Q^2 neutral-current DIS events close to the Born-level configuration $\gamma^* q \to q$ (Born kinematics) is presented. This cross section is measured deferentially as a function of the jet transverse momentum and pseudorapidity, as well as lepton-jet momentum imbalance and azimuthal angle correlation. The jets are reconstructed in the laboratory frame with the k_T algorithm and a distance parameter of 1.0. The data are corrected for detector effects using the OMNIFOLD method, which incorporates a simultaneous and unbinned unfolding in four dimensions using machine learning. The results are compared with leading order Mont Carlo event generators and higher order calculations performed within the context of collinear or transverse-momentum-dependent (TMD) factorization in Quantum Chromodynamics (QCD). The measurement probes a wide range of QCD phenomena, including TMD parton-distribution functions (PDFs) and their evolution with energy.

First application of OmniFold by an experimental collaboration!

In progress – OmniFolding jets in CMS Open Data to extract quark/gluon jet distributions

[PTK, Kryhin, Thaler, *to appear soon*]



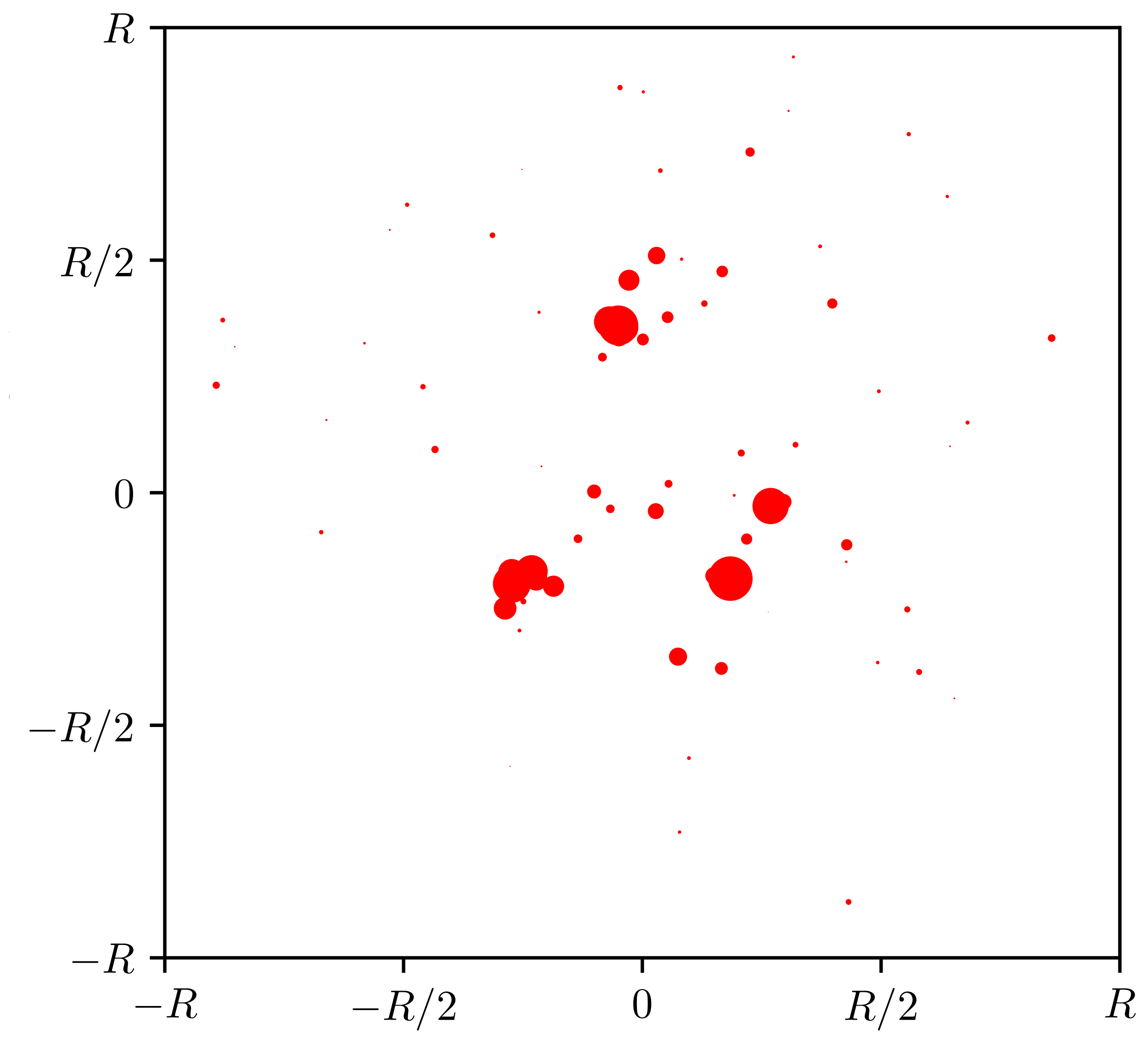
pronginess"
nginess"

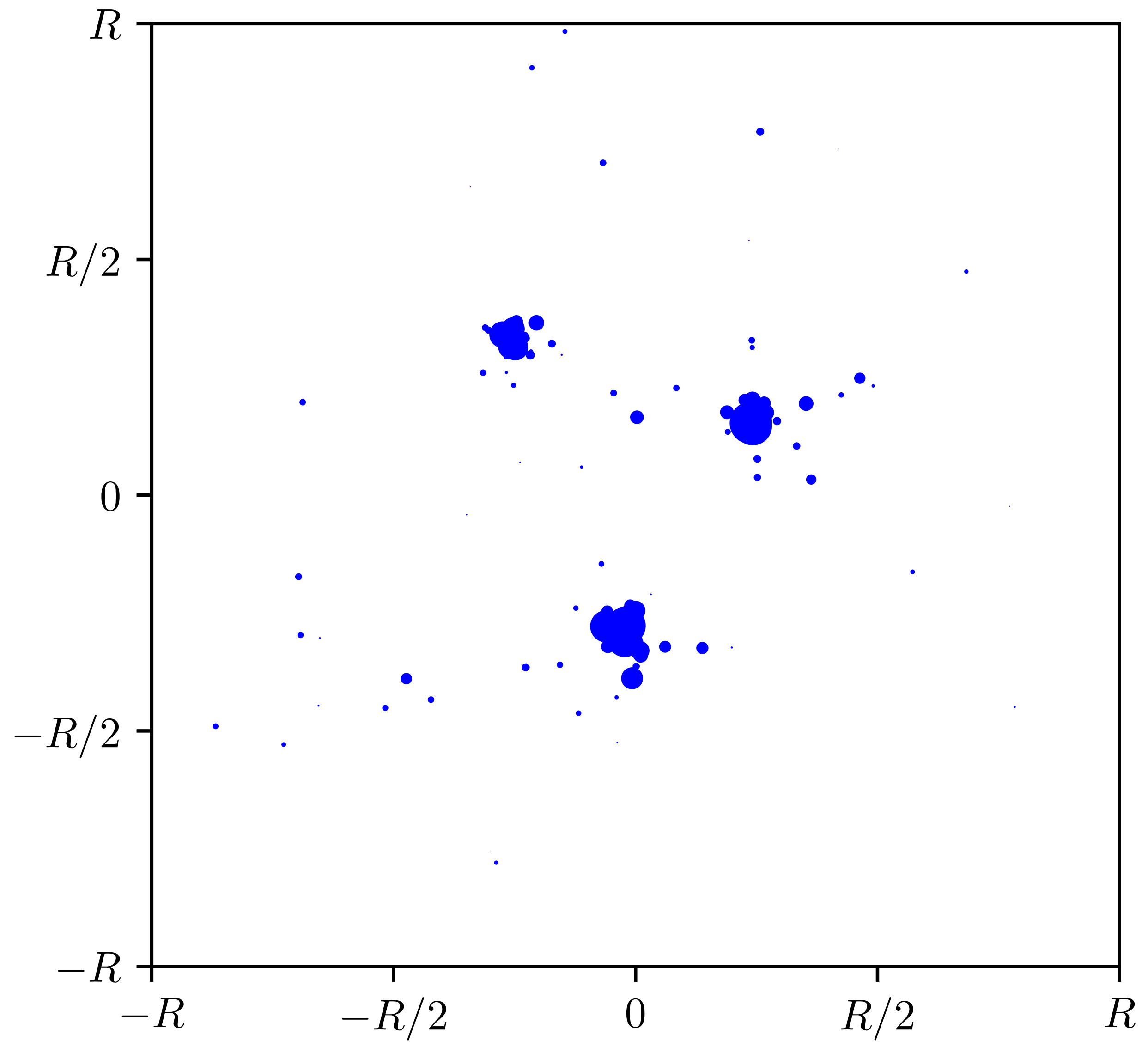
(Additional unfolded distributions in backup)

Optimal Transport in Particle Physics

When are Two Distributions similar?

Optimal transport minimizes the "work" (stuff x distance) required to transport supply to demand

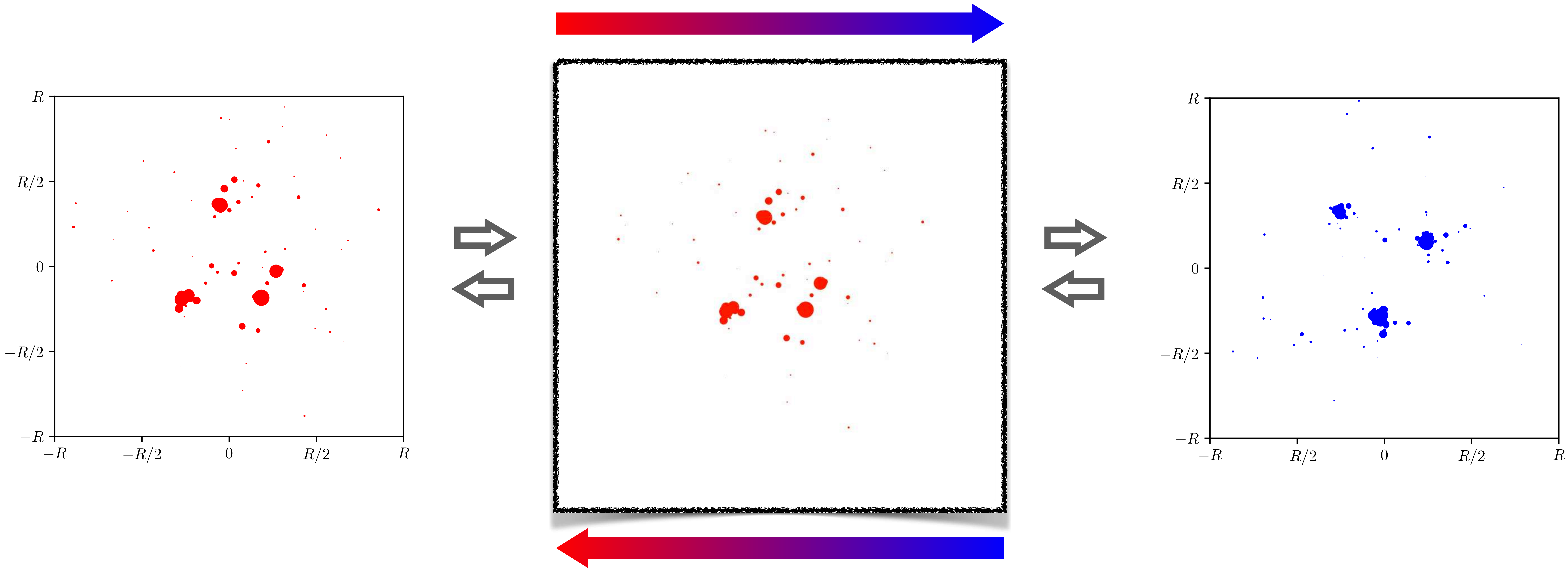




[Monge, 1781; Vaseršteĭn, 1969; Peleg, Werman, Rom, IEEE 1989; Rubner, Tomasi, Guibas, ICCV 1998, ICJV 2000; Pele, Werman, ECCV 2008; Pele, Taskar, GSI 2013]

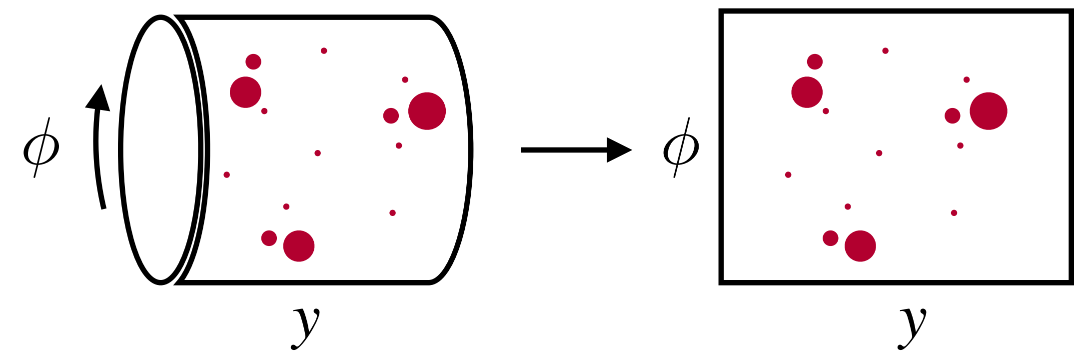
When are Two Distributions similar?

Optimal transport minimizes the "work" (stuff x distance) required to transport supply to demand

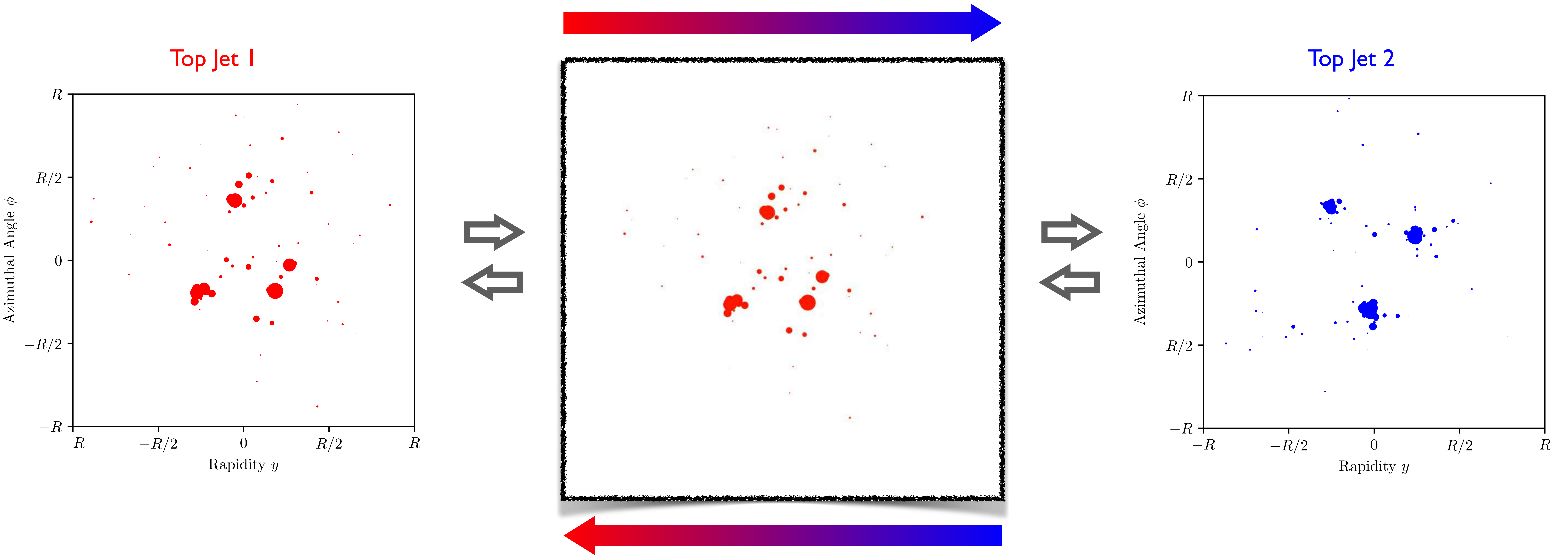


[Monge, 1781; Vaseršteĭn, 1969; Peleg, Werman, Rom, IEEE 1989;
Rubner, Tomasi, Guibas, ICCV 1998, ICJV 2000; Pele, Werman, ECCV 2008; Pele, Taskar, GSI 2013]

When are Two Distributions similar?

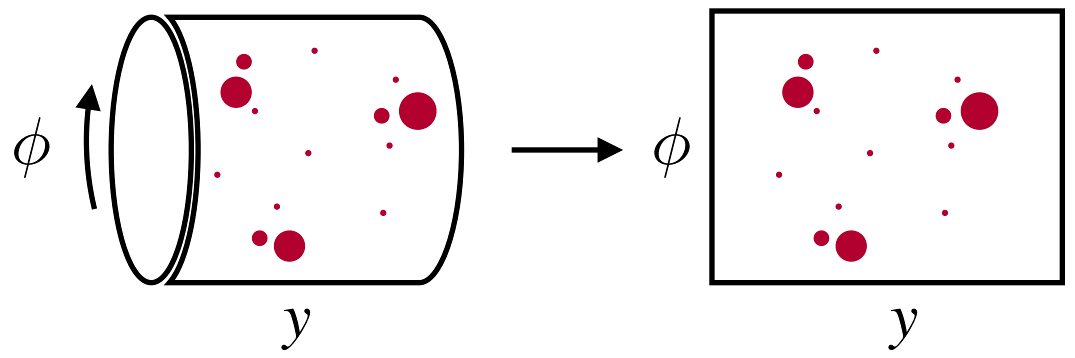


Optimal transport minimizes the "work" (stuff x distance) required to transport supply to demand



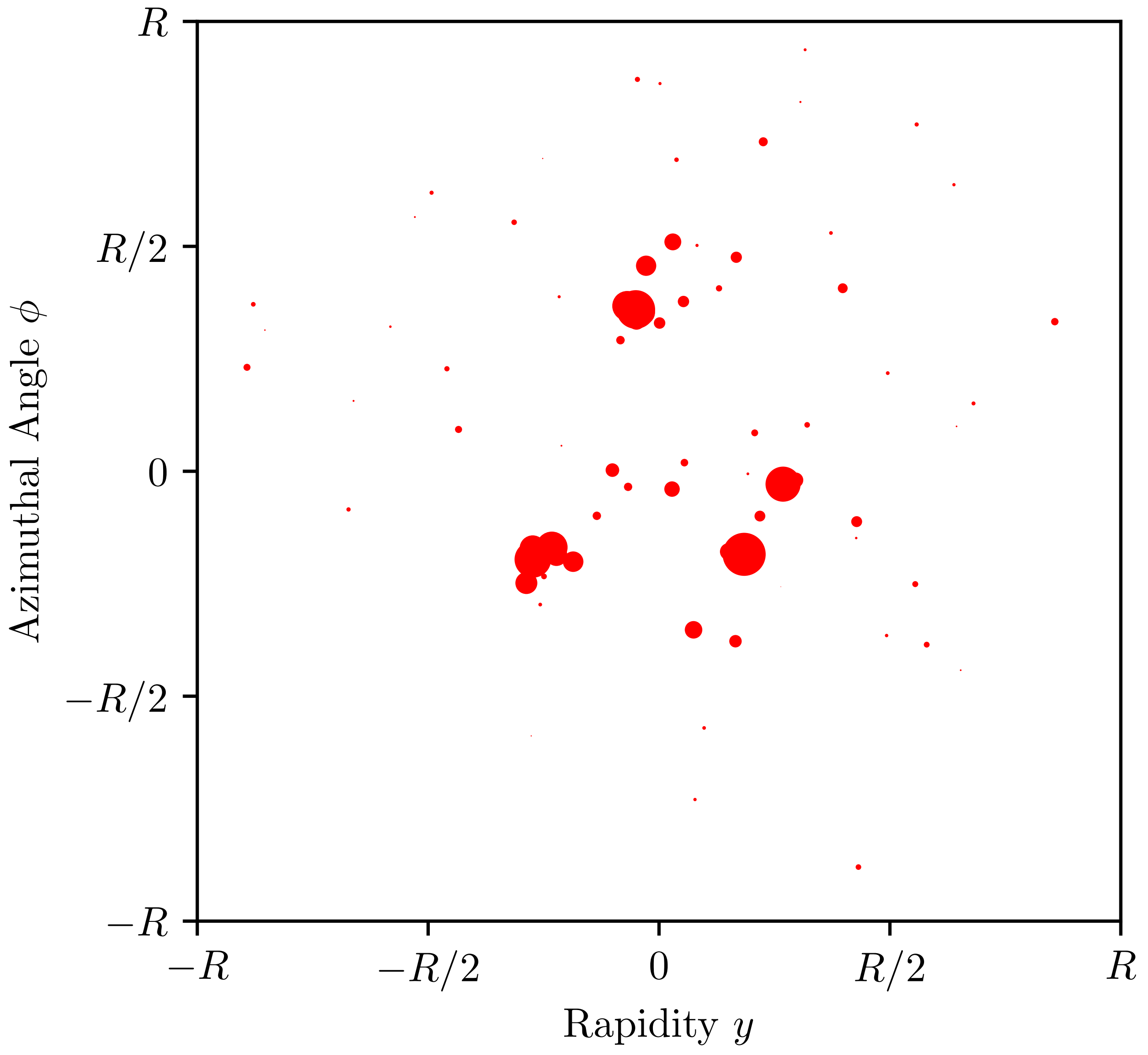
[Monge, 1781; Vaseršteĭn, 1969; Peleg, Werman, Rom, IEEE 1989;
Rubner, Tomasi, Guibas, ICCV 1998, ICJV 2000; Pele, Werman, ECCV 2008; Pele, Taskar, GSI 2013]

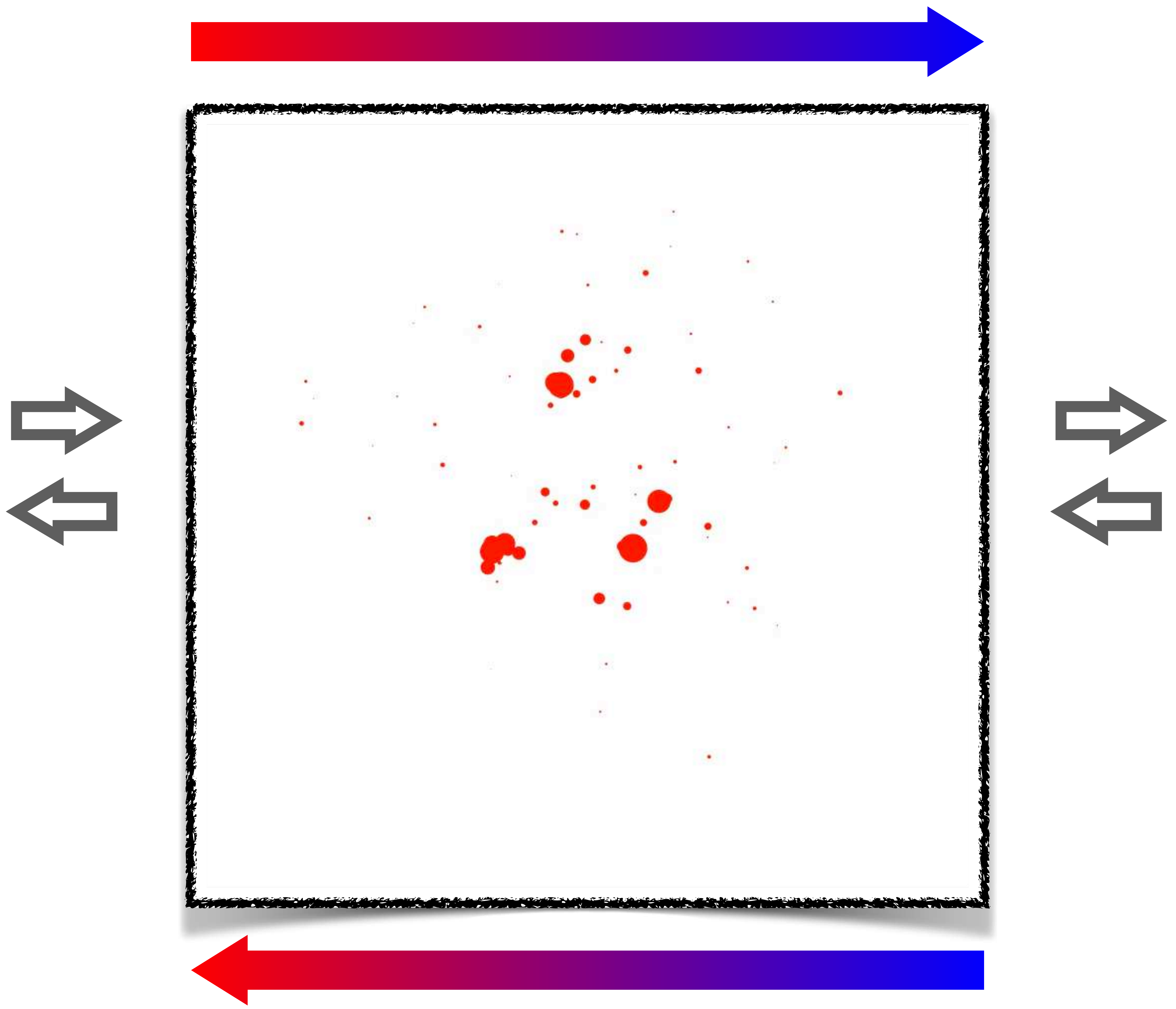
When are Two Distributions similar?



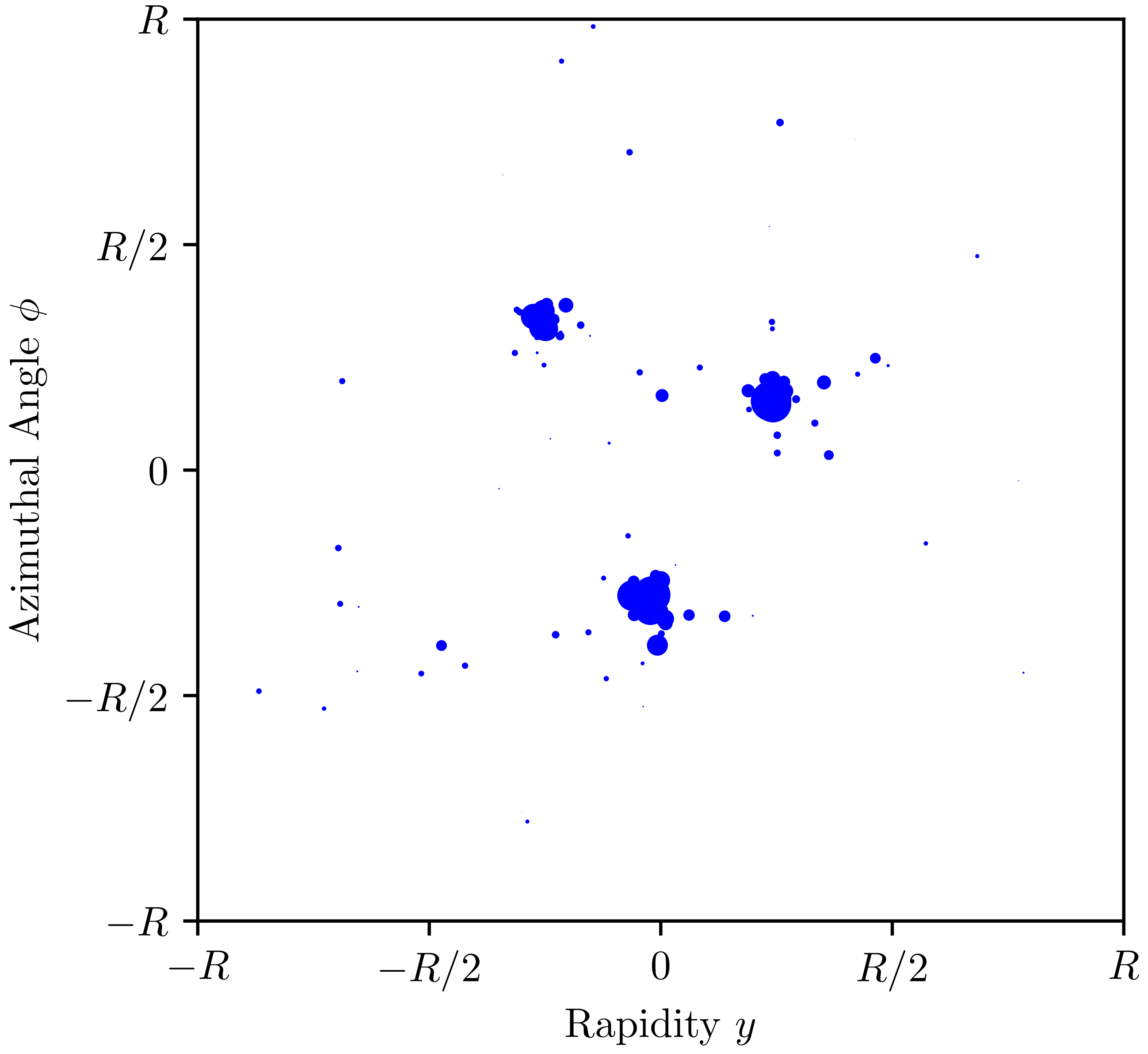
Optimal transport minimizes the "work" (stuff x distance) required to transport supply to demand

Top Jet 1





Top Jet 2



$$\mathcal{E}(\hat{n}) = \sum_{i=1}^{M} E_i \, \delta(\hat{n} - \hat{n}_i)$$

Energy Flow Distribution

$$\theta_{ij} = \sqrt{\Delta y_{ij}^2 + \Delta \phi_{ij}^2}$$

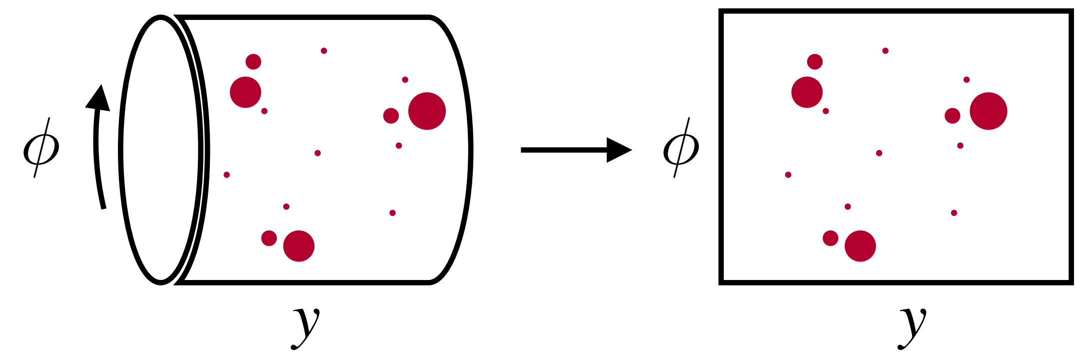
Ground Distance

Provides a metric on normalized distributions in a space with a ground distance measure

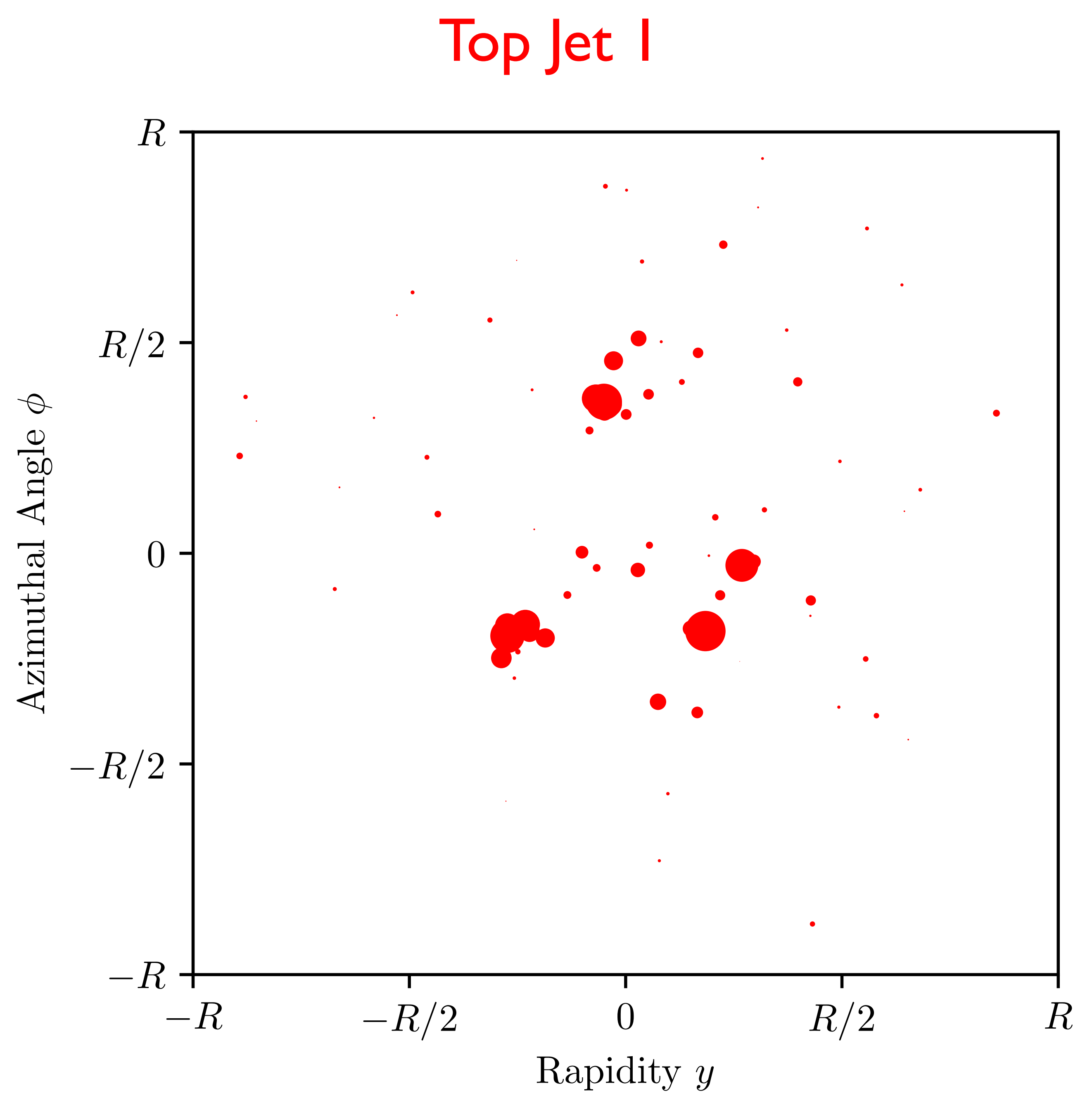
↳ symmetric, non-negative, triangle inequality, zero iff identical

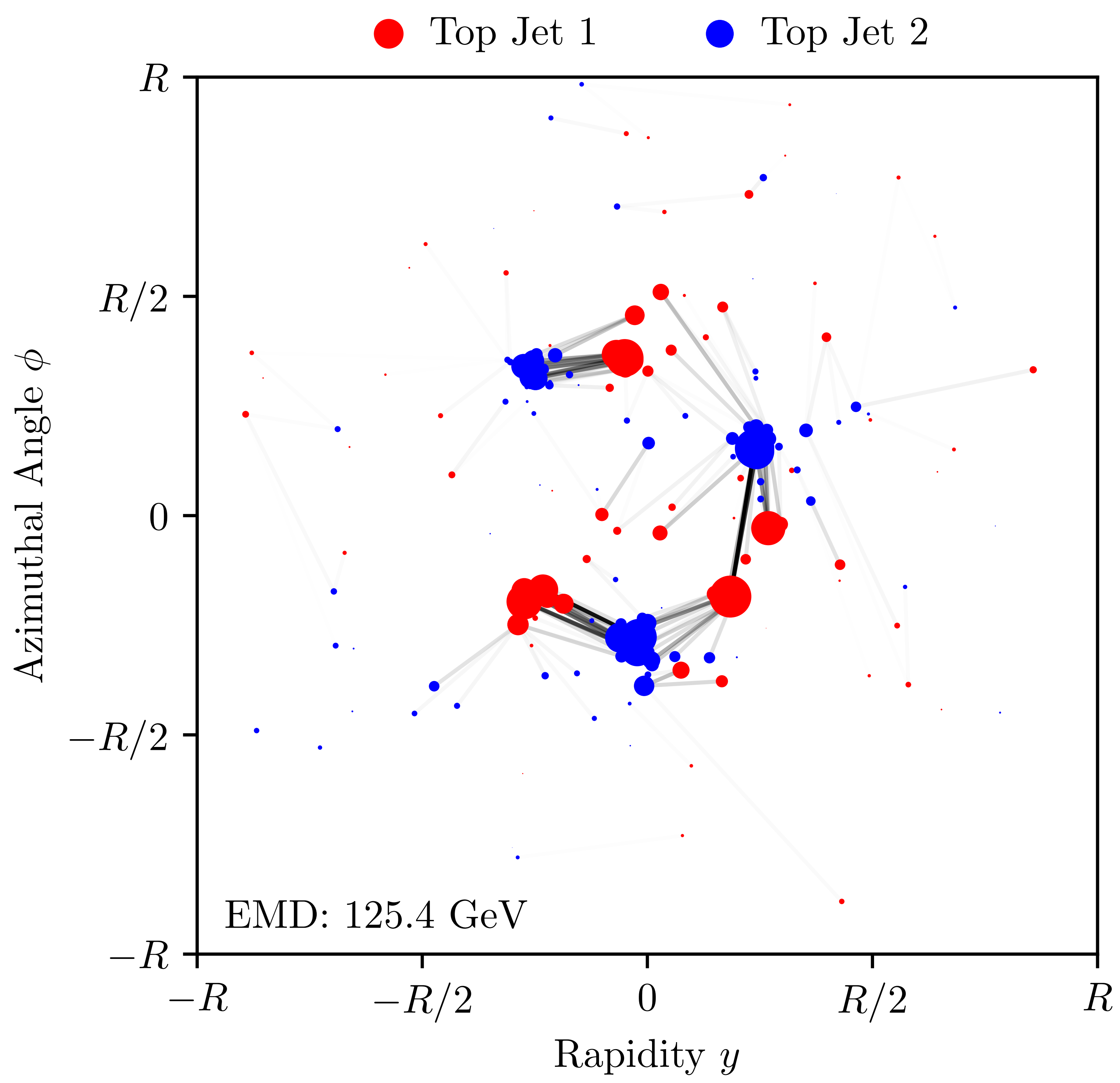
[Monge, 1781; Vaseršteĭn, 1969; Peleg, Werman, Rom, IEEE 1989; Rubner, Tomasi, Guibas, ICCV 1998, ICJV 2000; Pele, Werman, ECCV 2008; Pele, Taskar, GSI 2013]

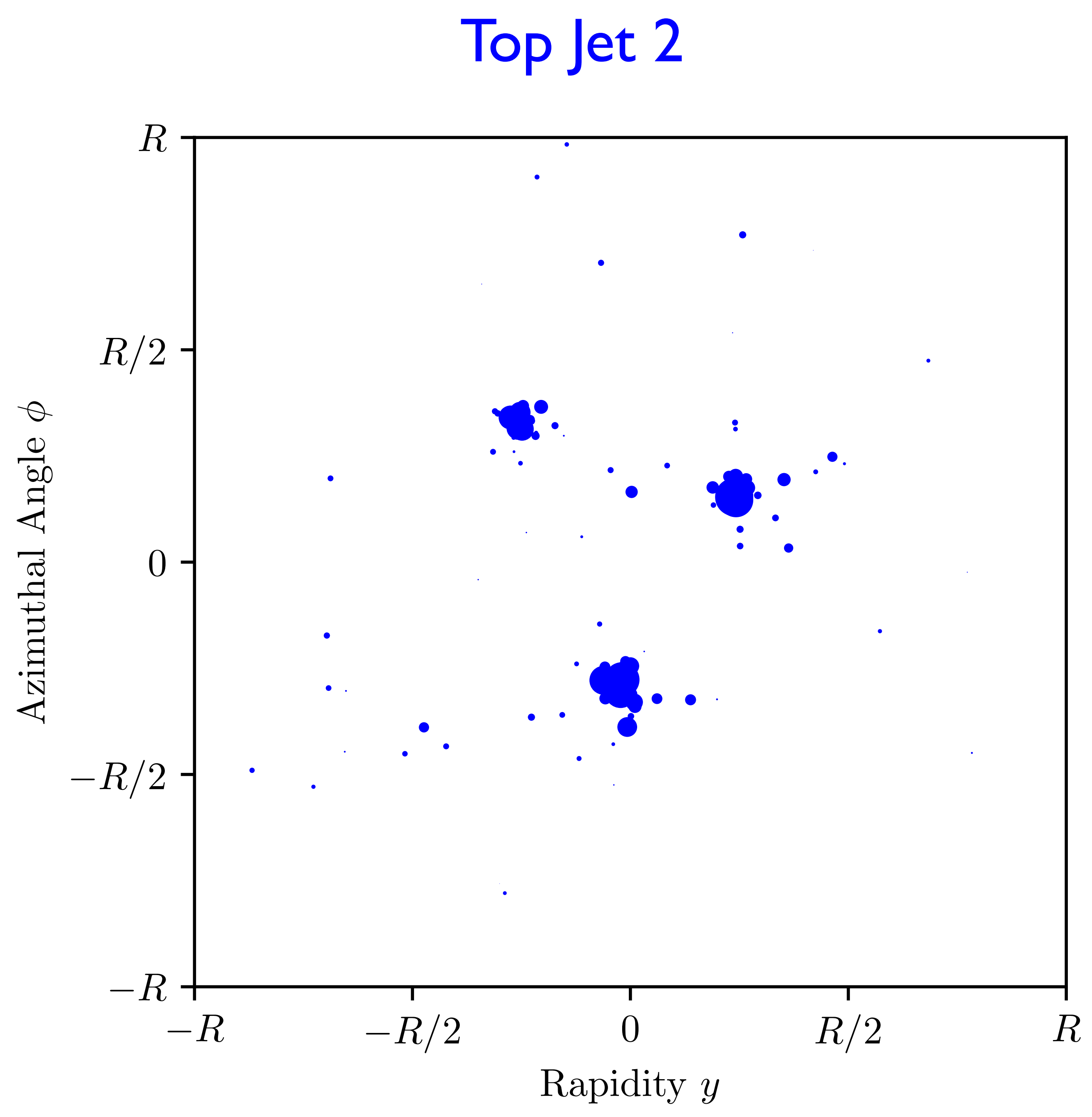
When are Two Distributions similar?



Optimal transport minimizes the "work" (stuff x distance) required to transport supply to demand







$$\mathcal{E}(\hat{n}) = \sum_{i=1}^{M} E_i \, \delta(\hat{n} - \hat{n}_i)$$

Energy Flow Distribution

Provides a metric on normalized distributions in a space with a ground distance measure

$$\theta_{ij} = \sqrt{\Delta y_{ij}^2 + \Delta \phi_{ij}^2}$$

Ground Distance

↳ symmetric, non-negative, triangle inequality, zero iff identical

[Monge, 1781; Vaseršteĭn, 1969; Peleg, Werman, Rom, IEEE 1989; Rubner, Tomasi, Guibas, ICCV 1998, IJCV 2000; Pele, Werman, ECCV 2008; Pele, Taskar, GSI 2013]

The Energy Mover's Distance (EMD)

[PTK, Metodiev, Thaler, PRL 2019 (editors' suggestion), Featured in Physics Magazine;
PTK, Metodiev, Thaler, JHEP 2020;
EnergyFlow and Wasserstein Python Packages]

EMD between energy flows defines a metric on the space of events

$$\mathrm{EMD}_{\beta,R}(\mathcal{E},\mathcal{E}') = \underbrace{\min_{\{f_{ij}\geq 0\}} \sum_i \sum_j f_{ij} \left(\frac{\theta_{ij}}{R}\right)^\beta}_{\text{Cost of optimal transport}} + \underbrace{\left|\sum_i E_i - \sum_j E'_j\right|}_{\text{Cost of energy creation}}$$

Distance (in GeV) — Cost of optimal transport — Cost of energy creation

$$\underbrace{\sum_j f_{ij} \leq E_i, \quad \sum_i f_{ij} \leq E'_j, \quad \sum_{ij} f_{ij} = \min\left(\sum_i E_i, \sum_j E'_j\right)}_{\text{Capacity constraints to ensure proper transport}}$$

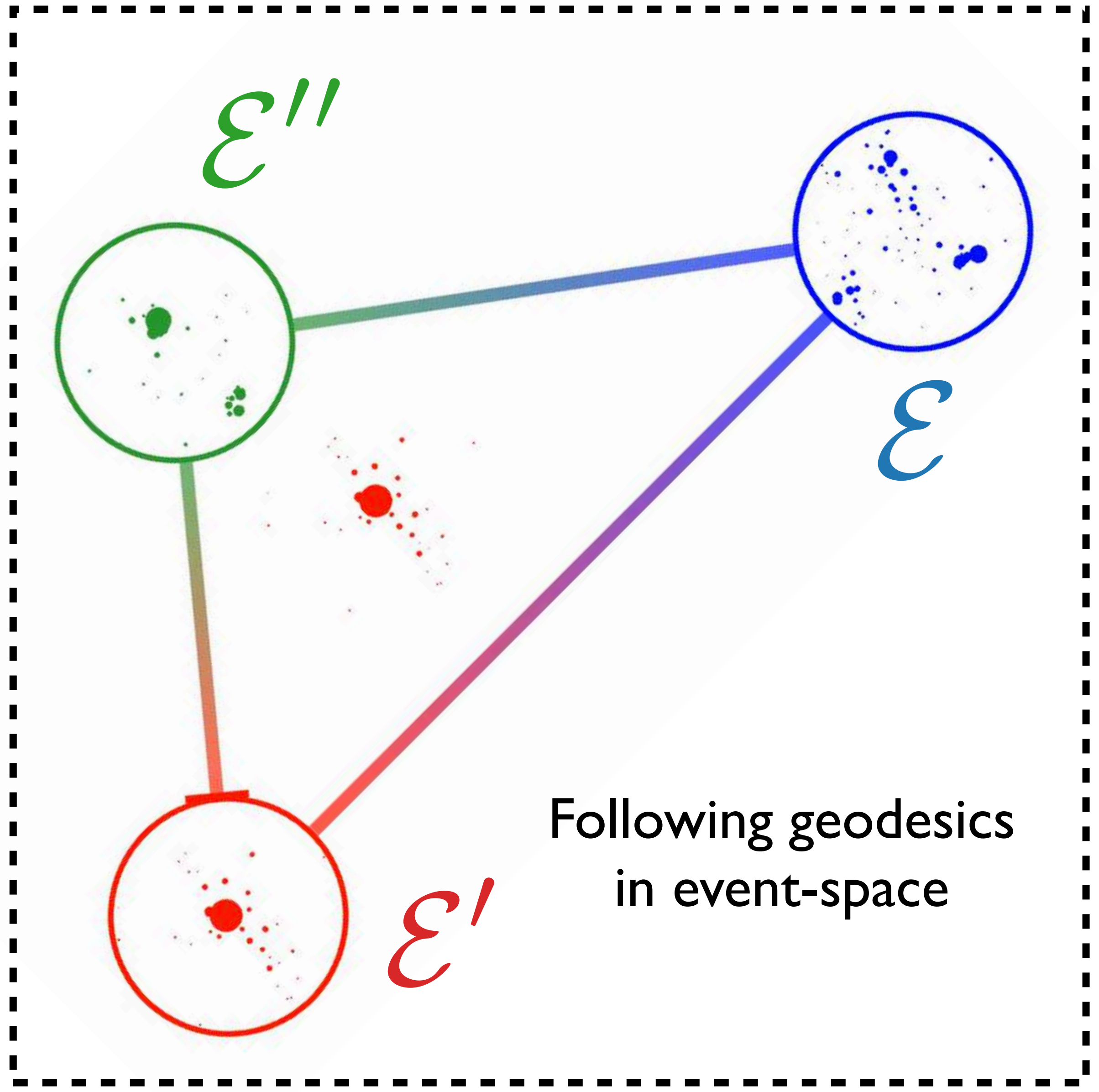
Capacity constraints to ensure proper transport

(θ_{ij}, f_{ij})

R: controls cost of transporting energy vs. destroying/creating it

β: angular weighting exponent

Abstract "space of events"



$$0 \leq \mathrm{EMD}(\mathcal{E},\mathcal{E}') \leq \mathrm{EMD}(\mathcal{E},\mathcal{E}'') + \mathrm{EMD}(\mathcal{E}'',\mathcal{E}')$$

Triangle inequality satisfied for $R \geq d_{\max}/2$

i.e. $R \geq$ jet radius for conical jets

The Energy Mover's Distance (EMD)

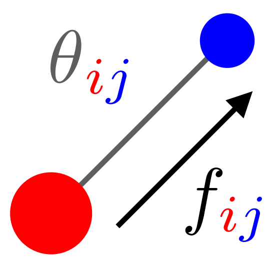
[PTK, Metodiev, Thaler, PRL 2019 (editors' suggestion), Featured in Physics Magazine;
PTK, Metodiev, Thaler, JHEP 2020;
EnergyFlow and Wasserstein Python Packages]

EMD between energy flows defines a metric on the space of events

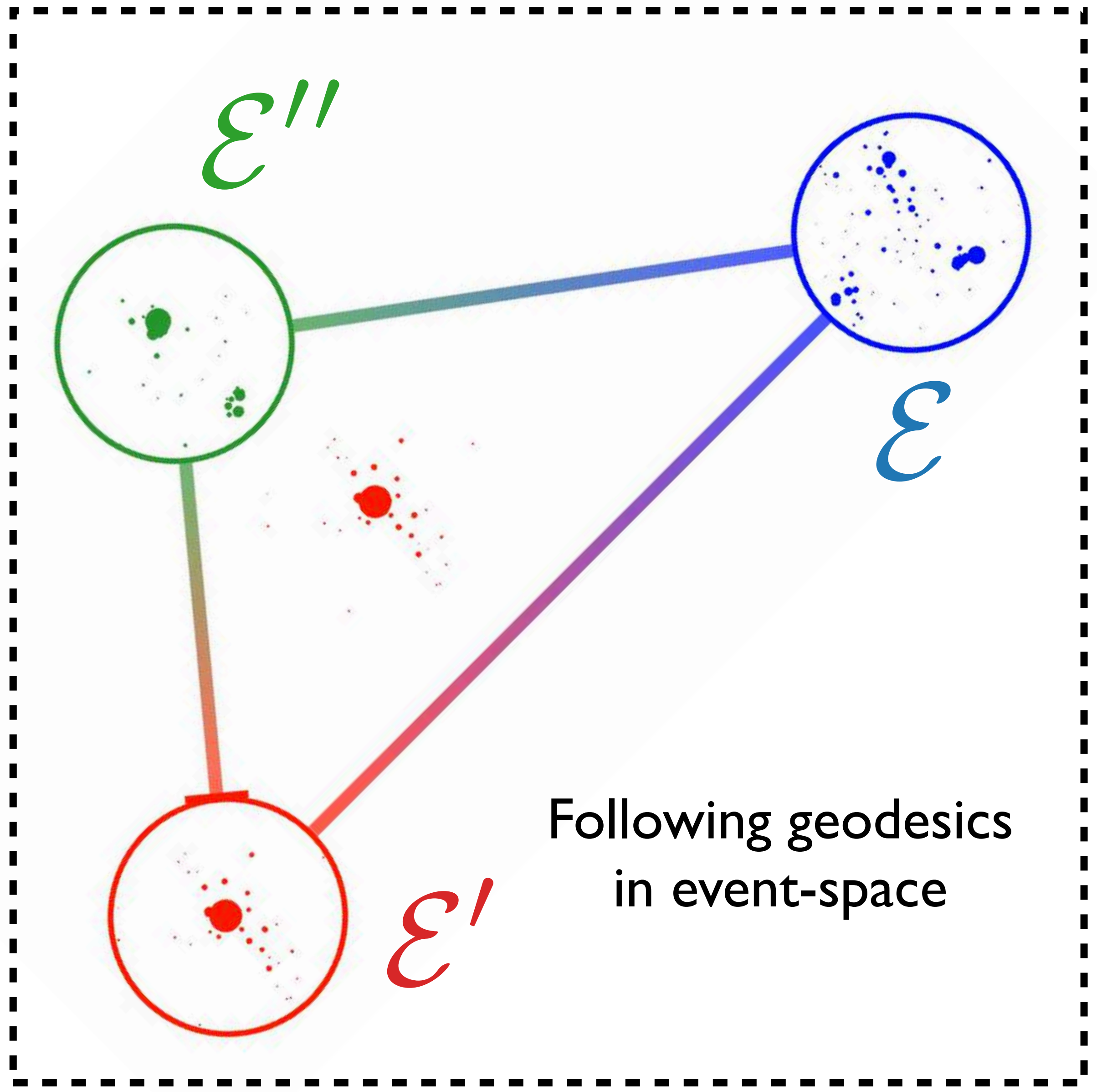
$$\mathrm{EMD}_{\beta,R}(\mathcal{E},\mathcal{E}') = \underbrace{\min_{\{f_{ij}\geq 0\}} \sum_i \sum_j f_{ij} \left(\frac{\theta_{ij}}{R}\right)^{\beta}}_{\text{Cost of optimal transport}} + \underbrace{\left| \sum_i E_i - \sum_j E'_j \right|}_{\text{Cost of energy creation}}$$

Distance (in GeV)

$$\underbrace{\sum_j f_{ij} \leq E_i, \quad \sum_i f_{ij} \leq E'_j, \quad \sum_{ij} f_{ij} = \min\left(\sum_i E_i, \sum_j E'_j\right)}_{\text{Capacity constraints to ensure proper transport}}$$



Abstract "space of events"



R: controls cost of transporting energy vs. destroying/creating it

β: angular weighting exponent

$$0 \leq \mathrm{EMD}(\mathcal{E},\mathcal{E}') \leq \mathrm{EMD}(\mathcal{E},\mathcal{E}'') + \mathrm{EMD}(\mathcal{E}'',\mathcal{E}')$$

Triangle inequality satisfied for $R \geq d_{\max}/2$

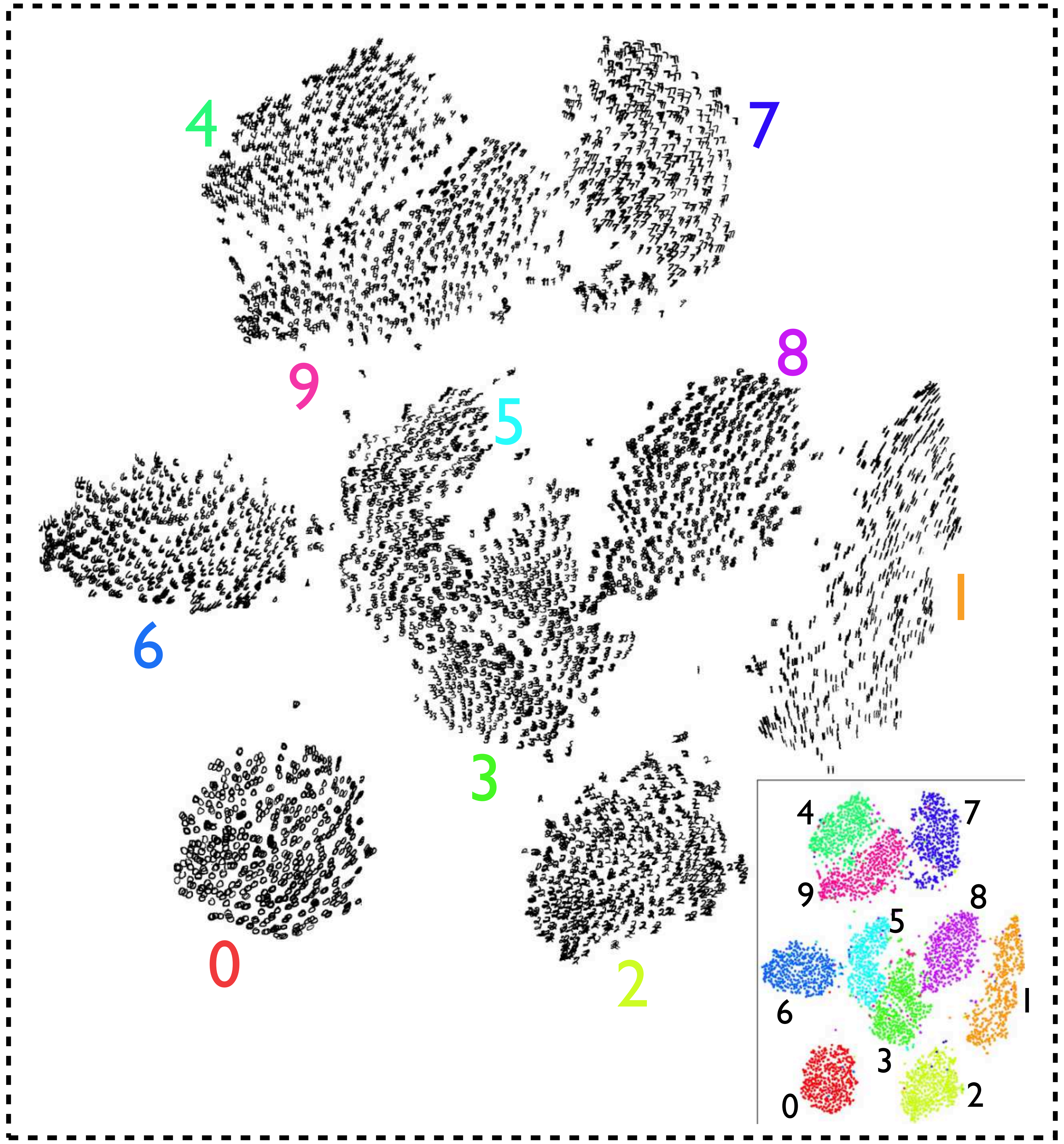
i.e. $R \geq$ jet radius for conical jets

Visualizing Geometry in the Space of Events

[PTK, Mastandrea, Metodiev, Naik, Thaler, PRD 2019;
code and datasets at energyflow.network]

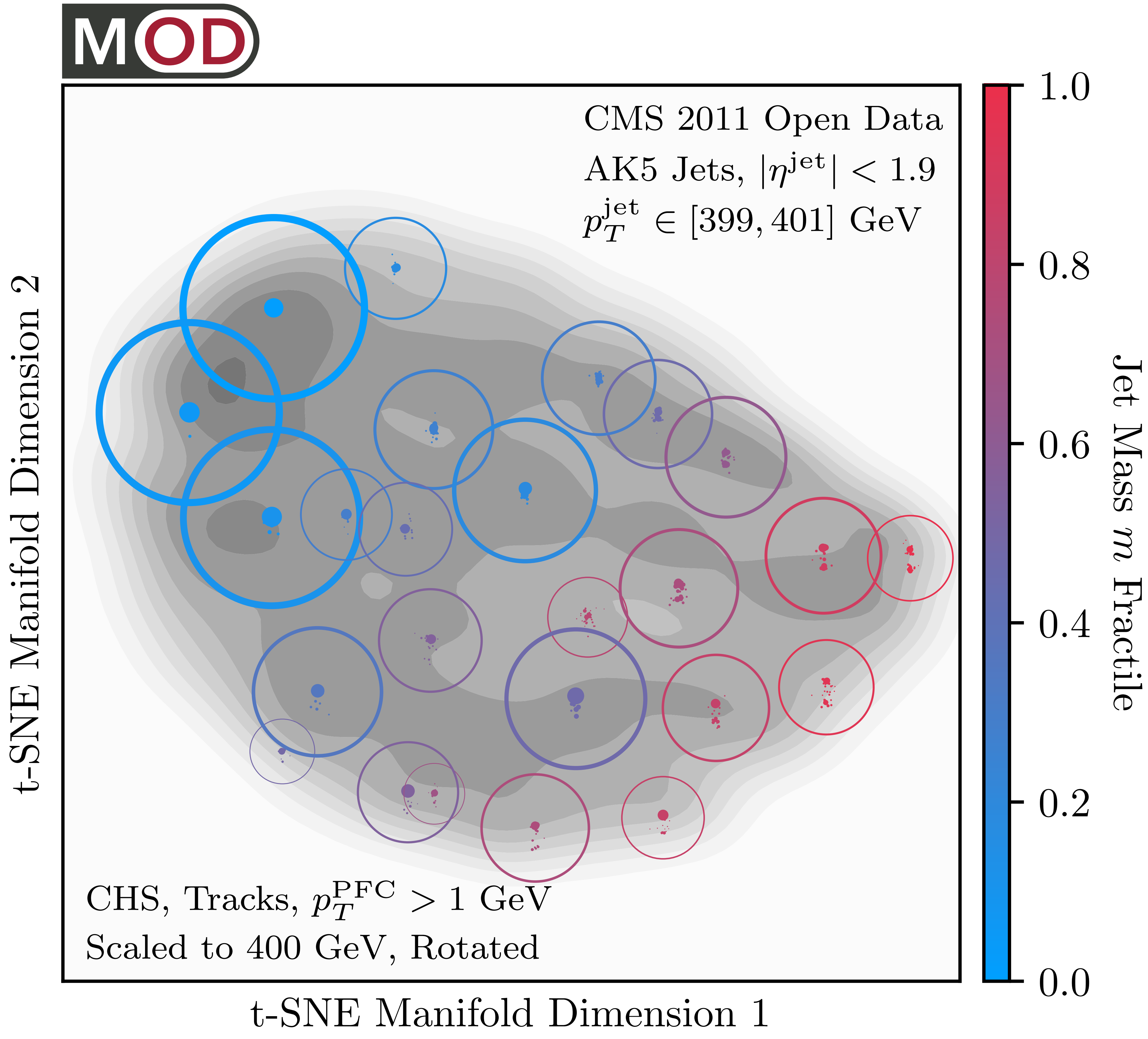
t-Distributed Stochastic Neighbor Embedding (t-SNE)

MNIST handwritten digits



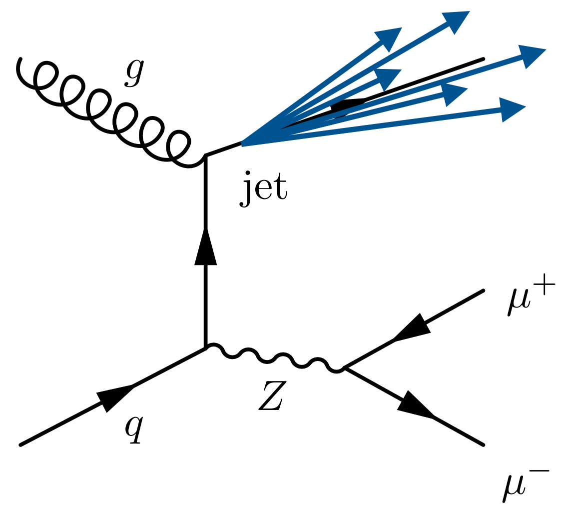
[L. van der Maaten, G. Hinton, JMLR 2008]

Jets from the CMS 2011 Open Data

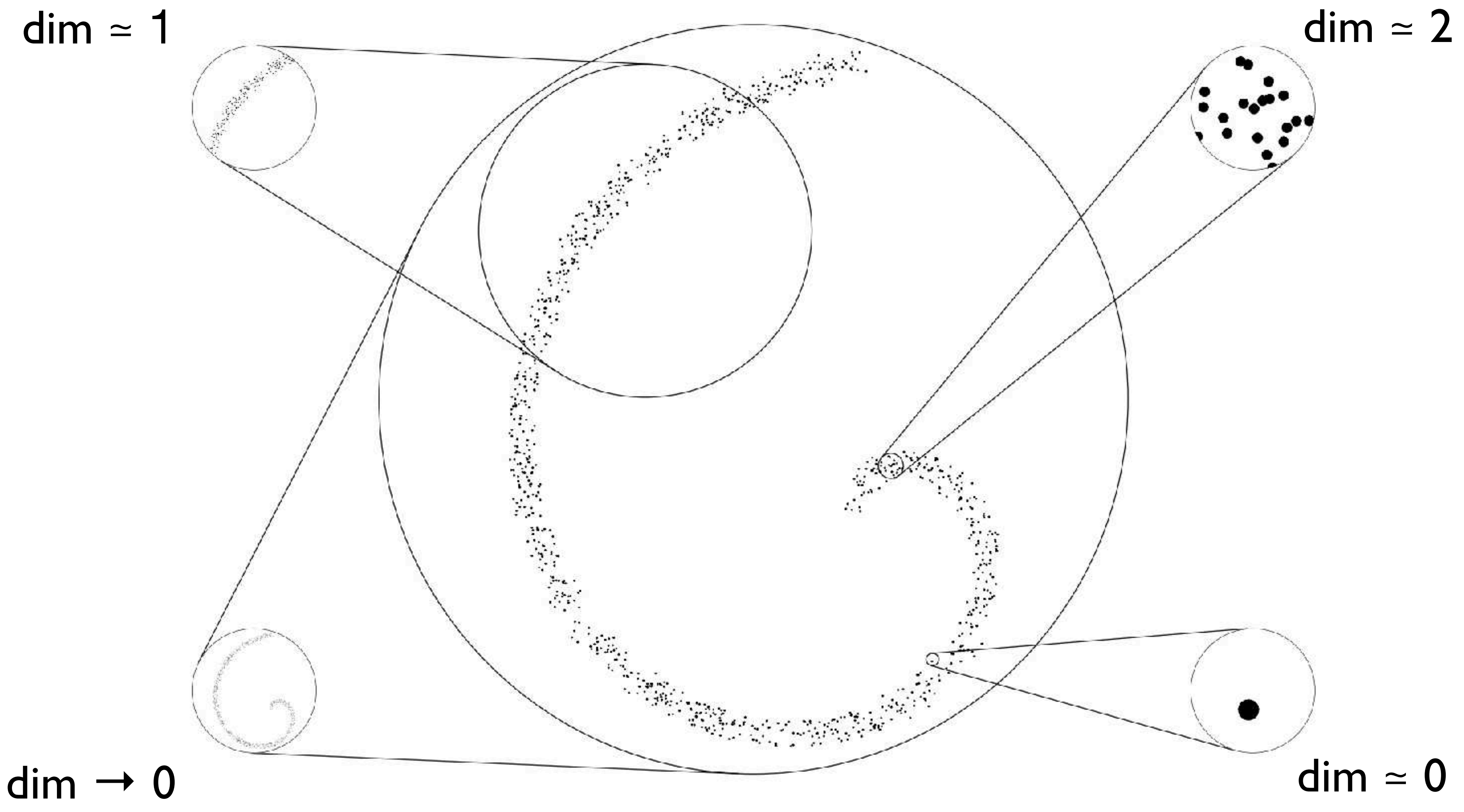


25 most representative jets ("medoids")
Size is proportional to cross section associated to that medoid

Unfolding Beyond Observables



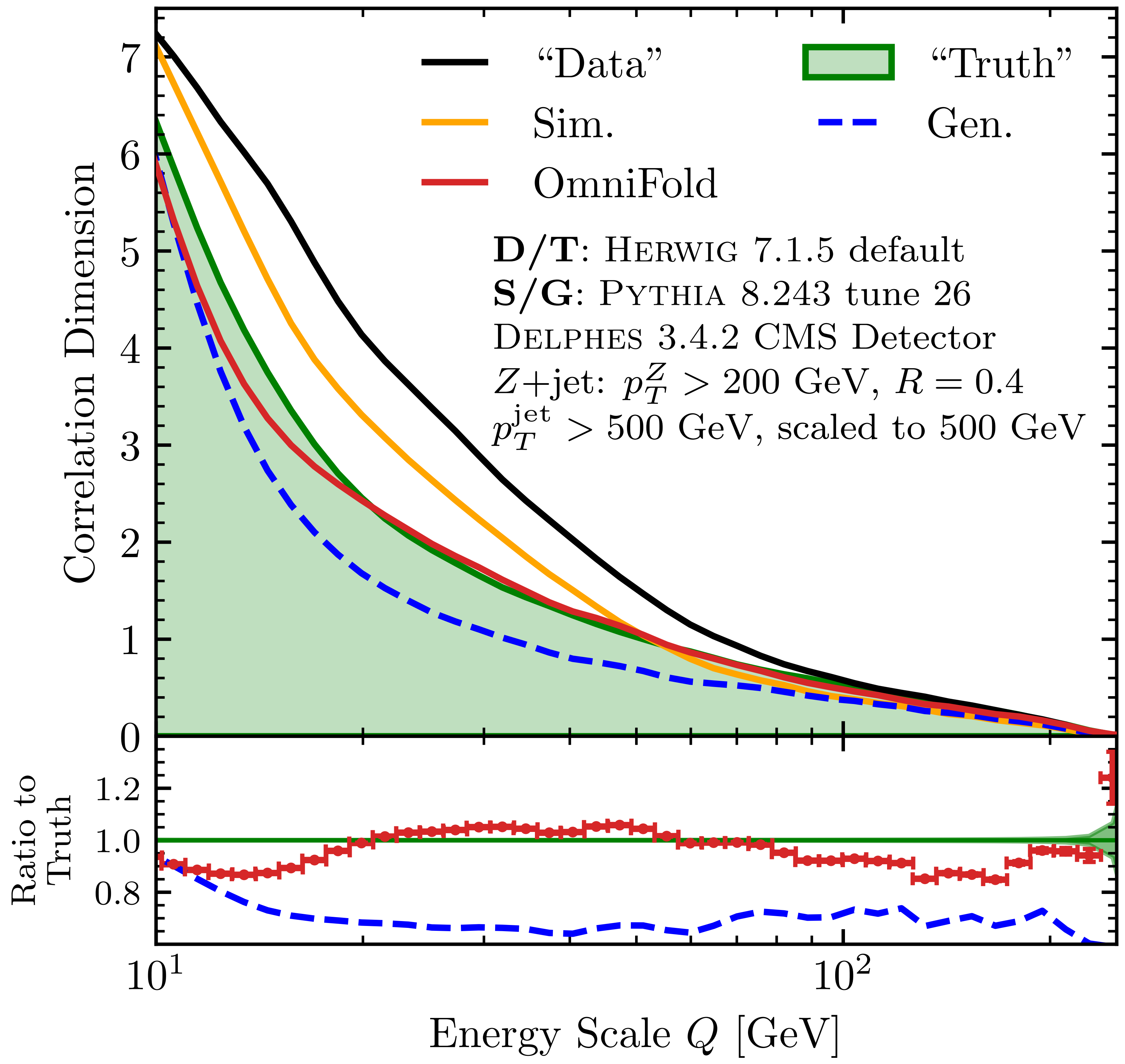
Correlation dimension: *how does the # of elements within a ball of size Q change?*



$$N_{\text{neigh.}}(Q) \propto Q^{\text{dim}} \implies \dim(Q) = Q\frac{d}{dQ}\ln N_{\text{neigh.}}(Q)$$

$$\dim(Q) = Q\frac{\partial}{\partial Q}\ln\sum_i\sum_j \boxed{w_i w'_j}\,\Theta(\text{EMD}(\mathcal{E}_i,\mathcal{E}'_j) < Q)$$

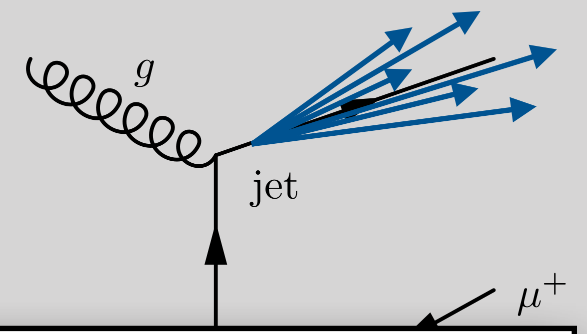
Weighted events naturally accommodated



Same OmniFold training can unfold a complicated function of pairs of events!

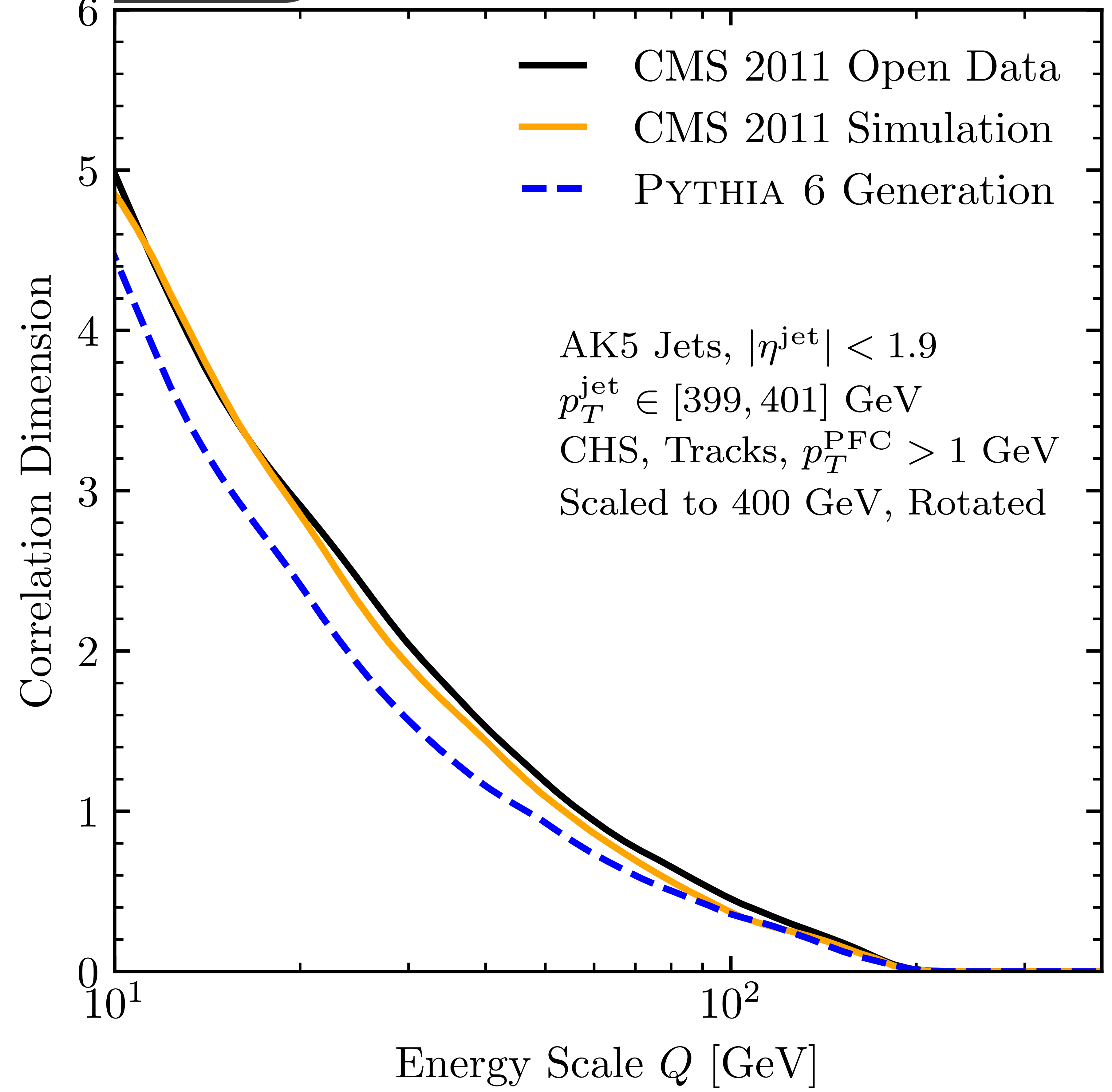
Larger detector effects and loss of stats seen at low Q

Unfolding Beyond Observables

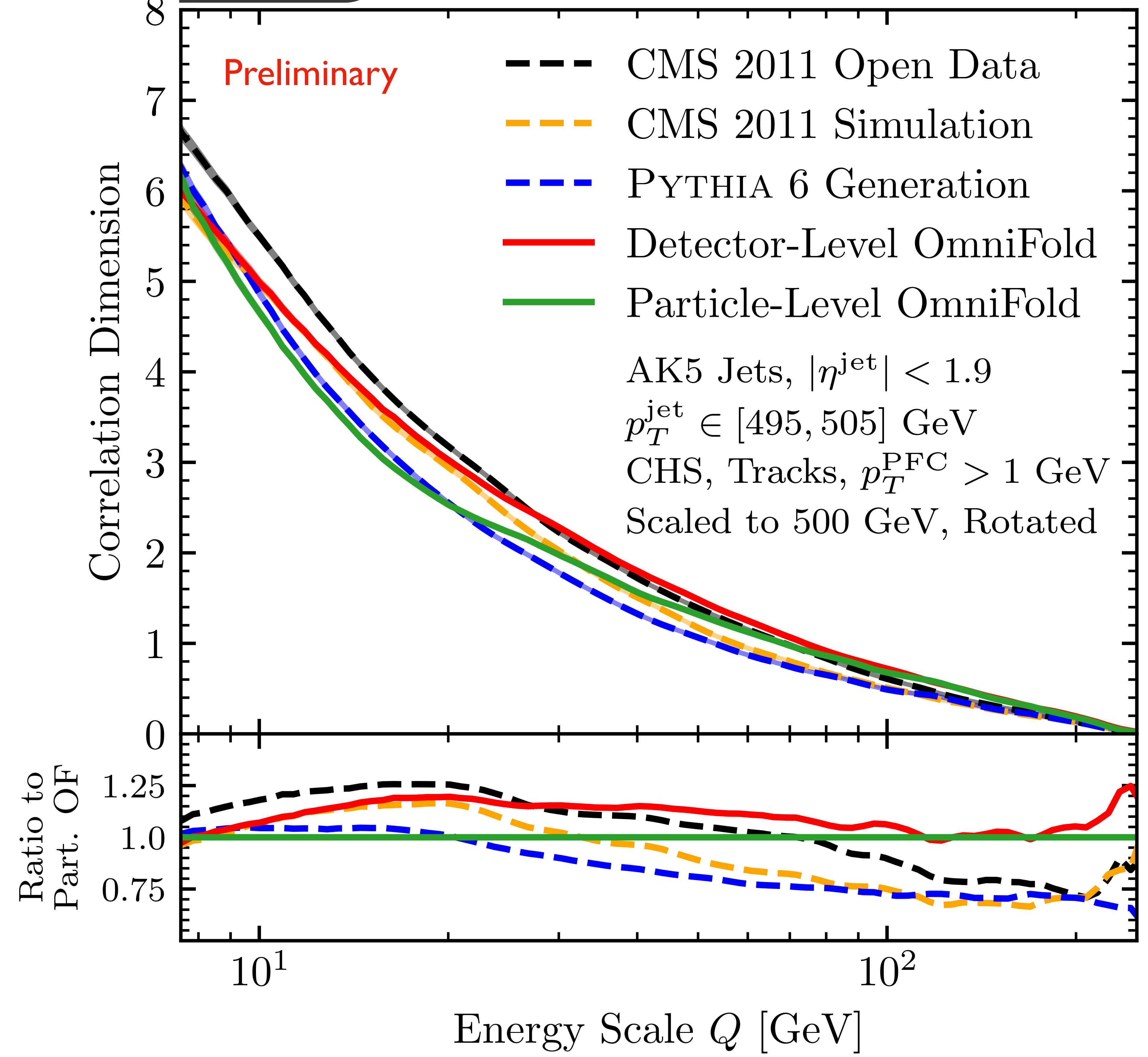


… in CMS Open Data

[PTK, Mastandrea, Metodiev, Naik, Thaler, PRD 2019]



[PTK, Kryhin, Thaler, *to appear soon*]



Particle Physics Fundamentals – Jets

Jets are critical to the success of the modern collider program

Architectures for Colliders – EFNs/PFNs

Simple, extensible neural network architecture(s) for collider events

Statistical Deconvolution – OmniFold

Likelihood-free inference uses high-dimensional classifiers (PFNs) to avoid explicit histograms and overcome the curse of dimensionality in unfolding

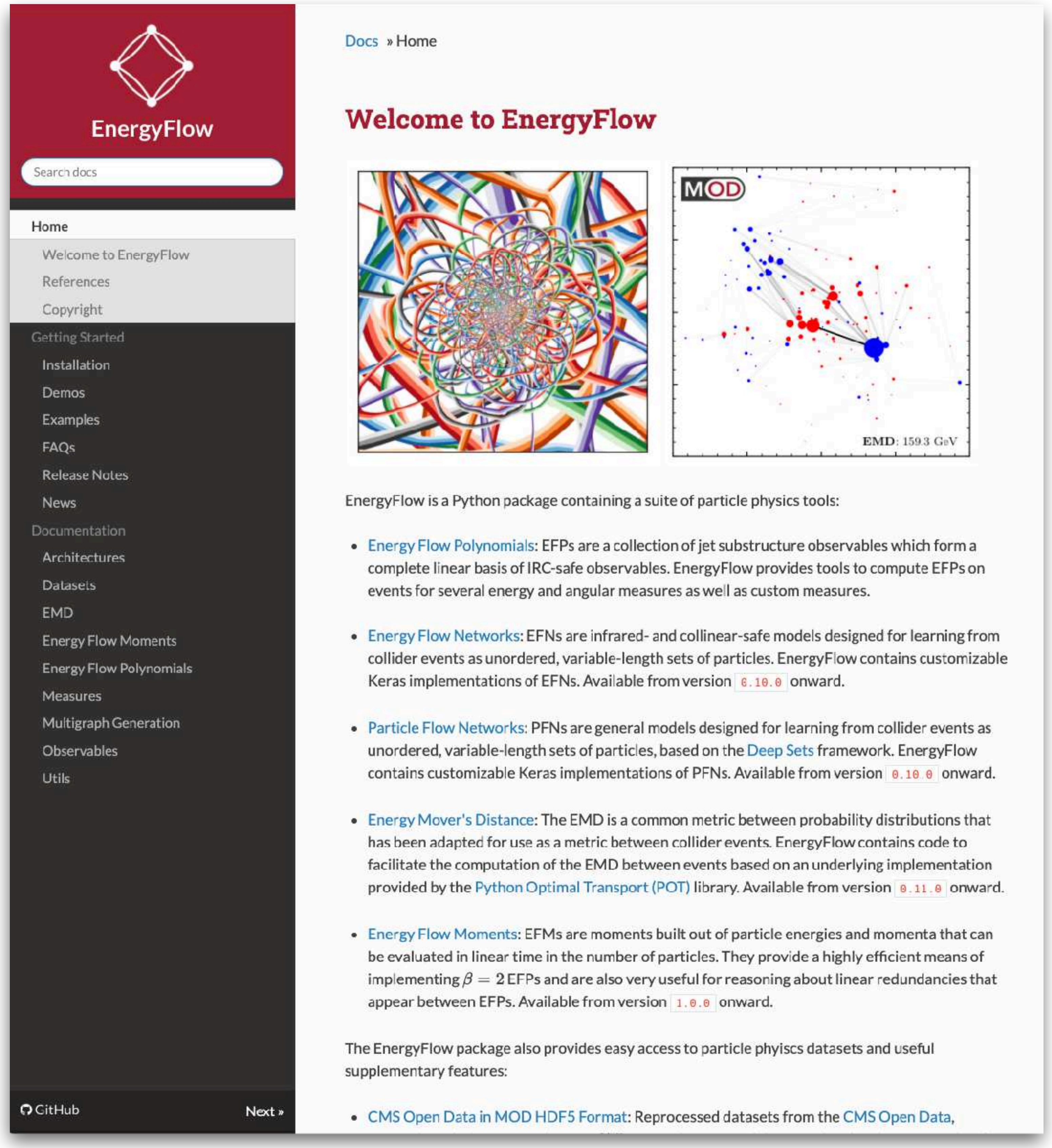
EnergyFlow Python Package

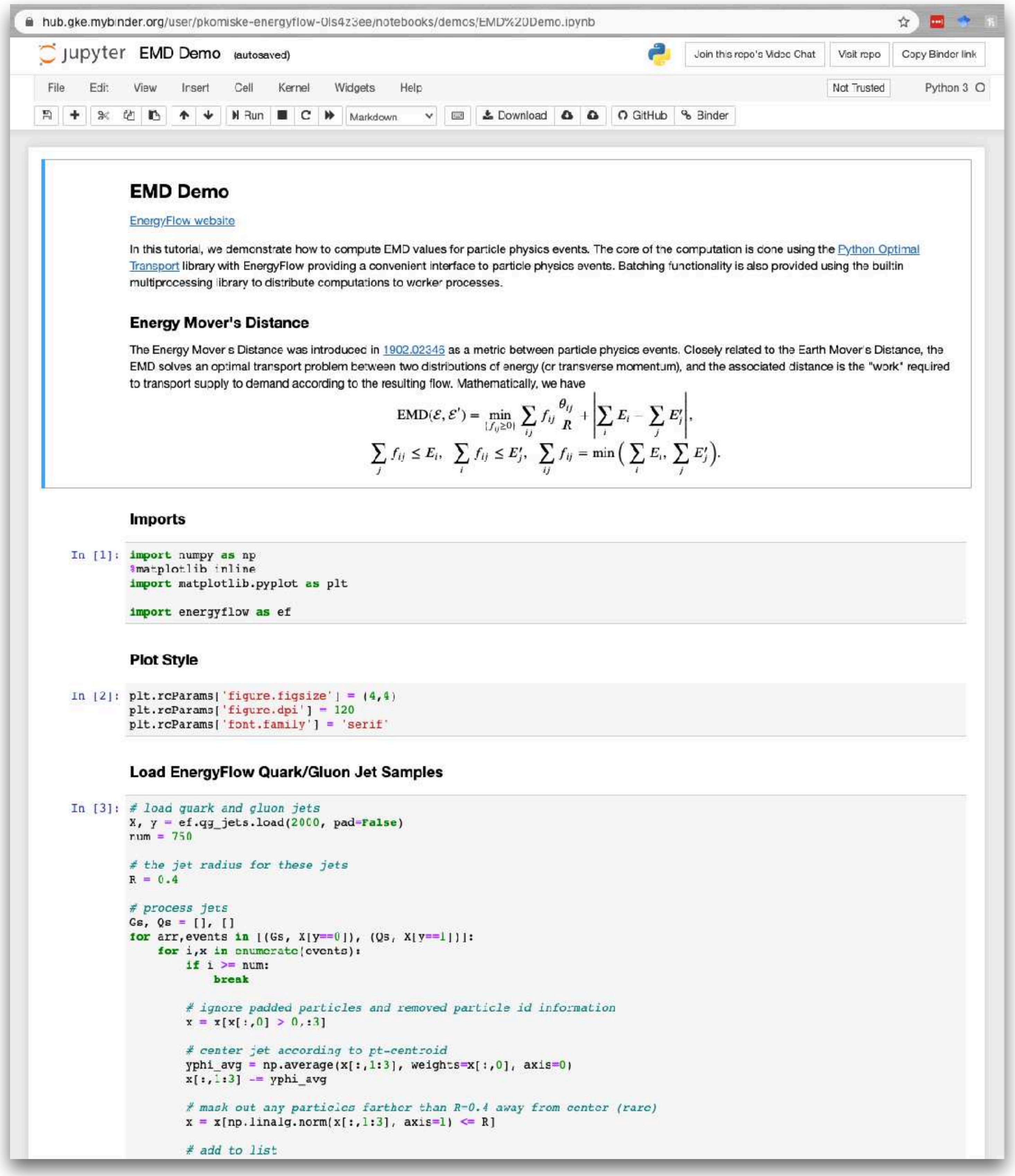
`pip3 install energyflow`

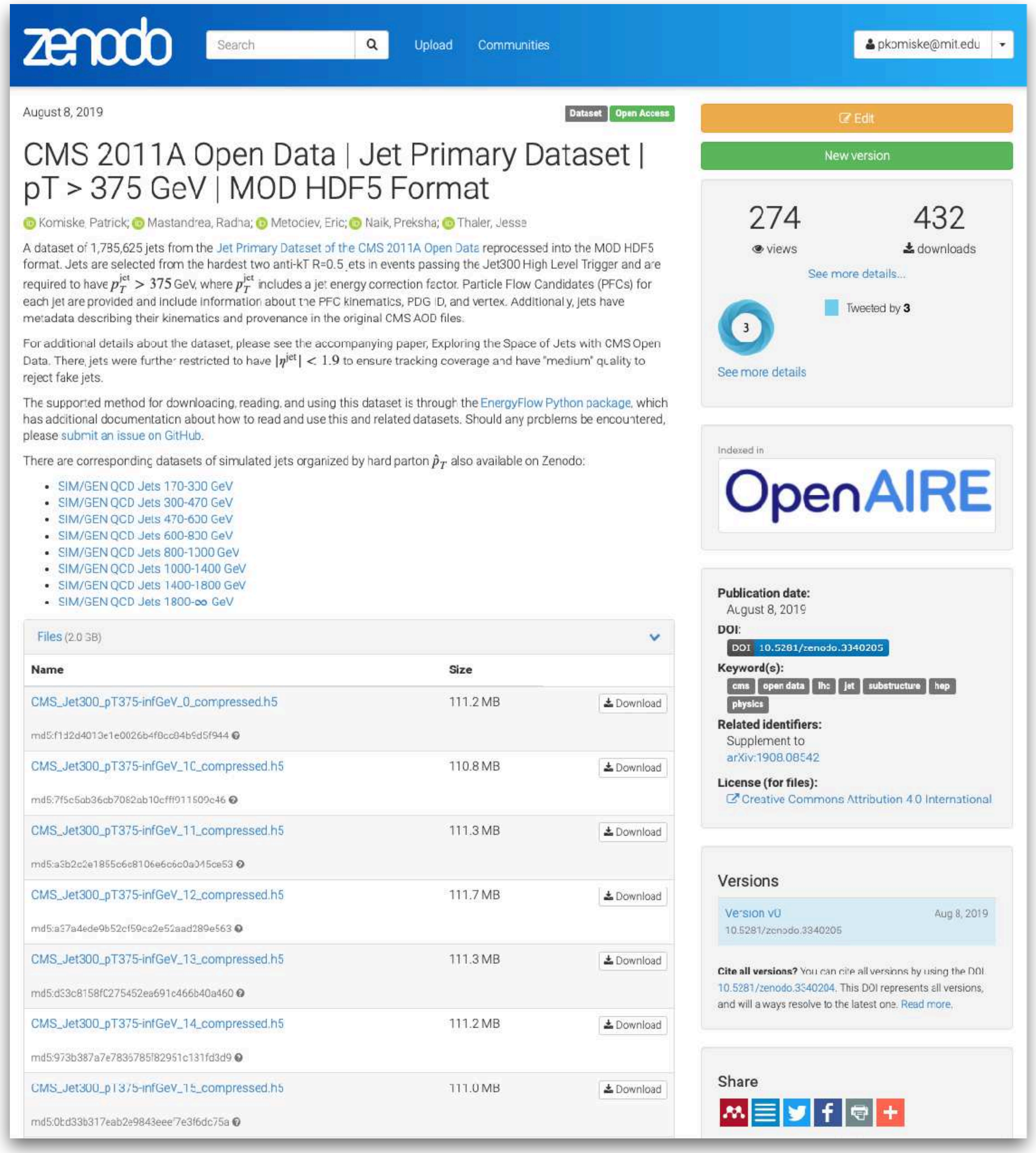
Implementations of EFNs/PFNs in Tensorflow, parallelized EMD calculations (in C++)

Detailed examples, demos, and documentation

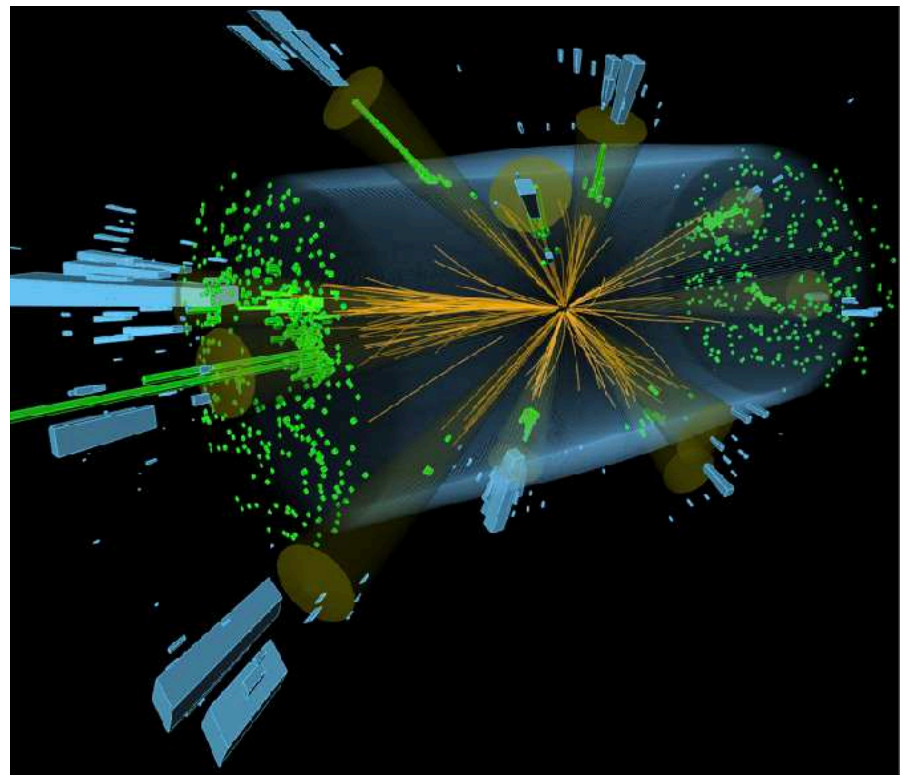
Interfaces with CMS 2011A Jet Primary Dataset (and other datasets) hosted on Zenodo





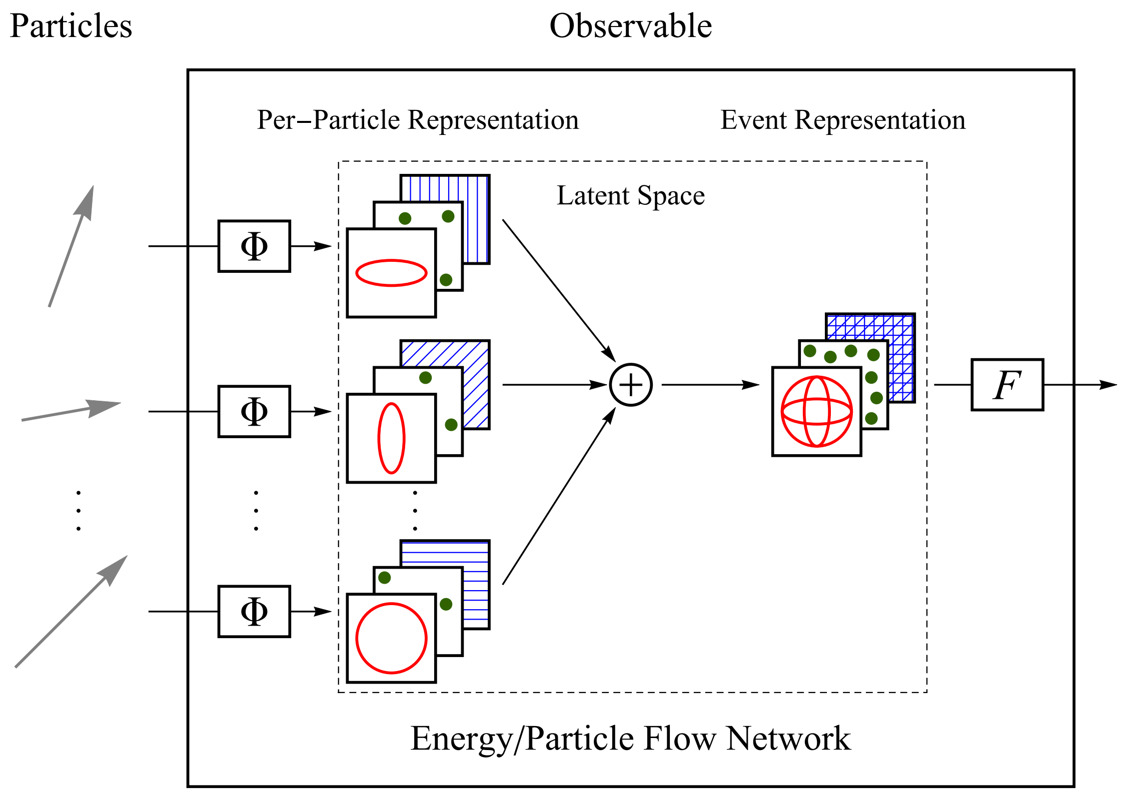


https://energyflow.network



Particle Physics Fundamentals – Jets

Jets and jet substructure will be essential to the next big collider discovery



Architectures for Colliders – EFNs/PFNs

EFNs/PFNs enable simple, fast, powerful deployment of deep learning for high-energy collider events

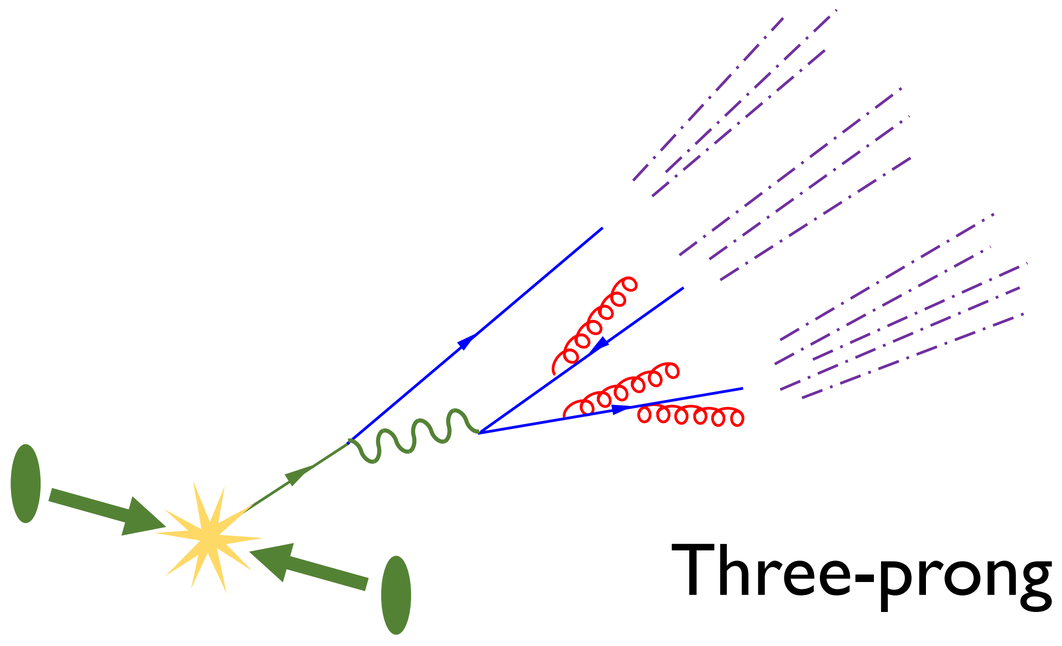
Statistical Deconvolution – OmniFold

Substantially improved unfolding enables multi-differential measurements with smaller uncertainties

Thank you!

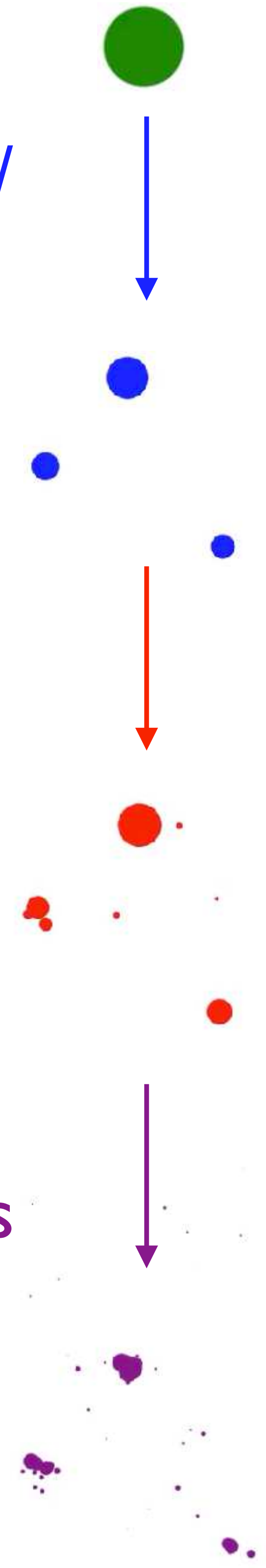
Additional Slides

Visualizing Jet Formation – Top Jets



Three-prong decay

500 GeV top quark



decay to W (decays to quark/ anti-quark) and a b quark

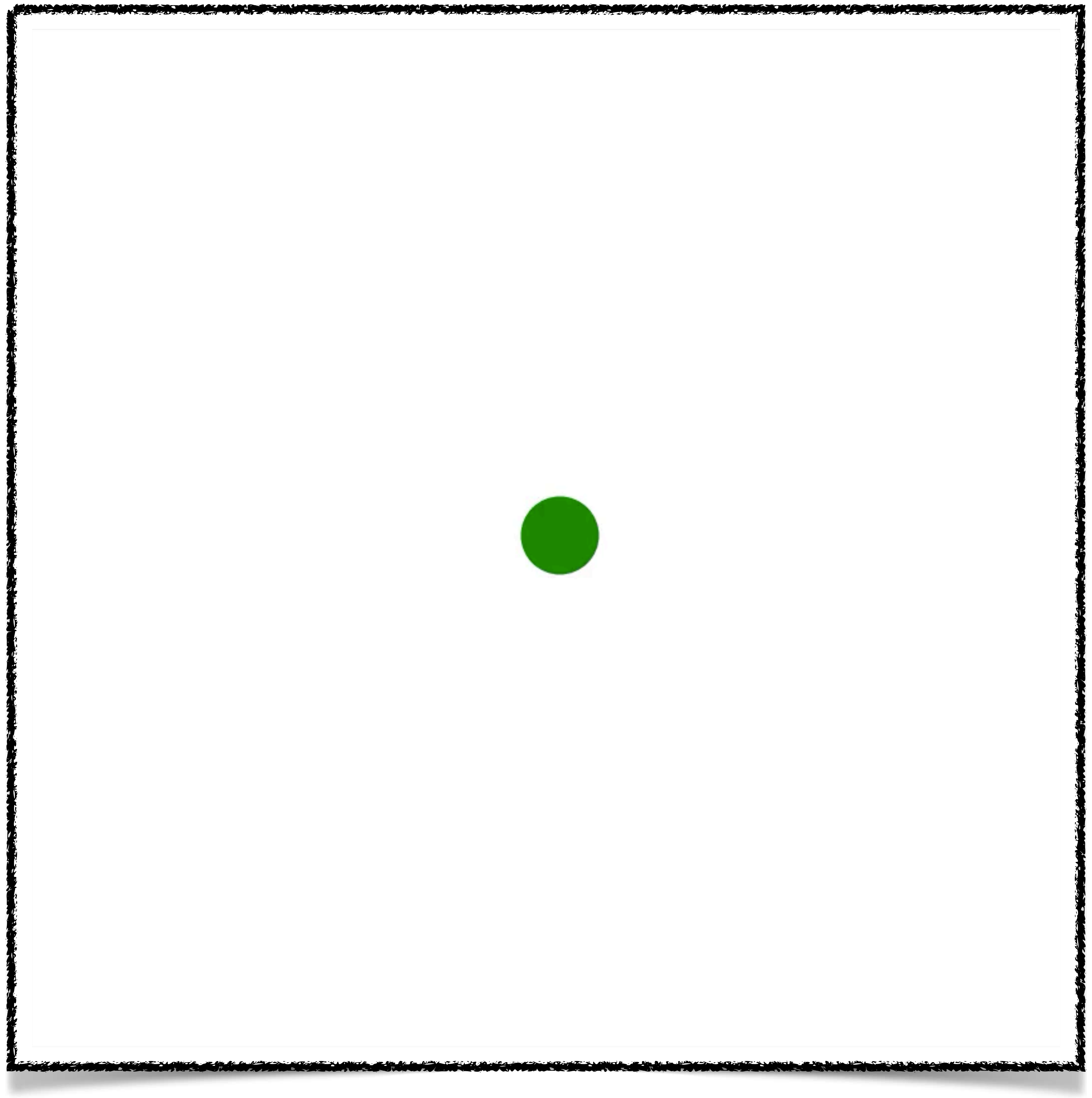
161.1 GeV

fragmentation into quarks and gluons

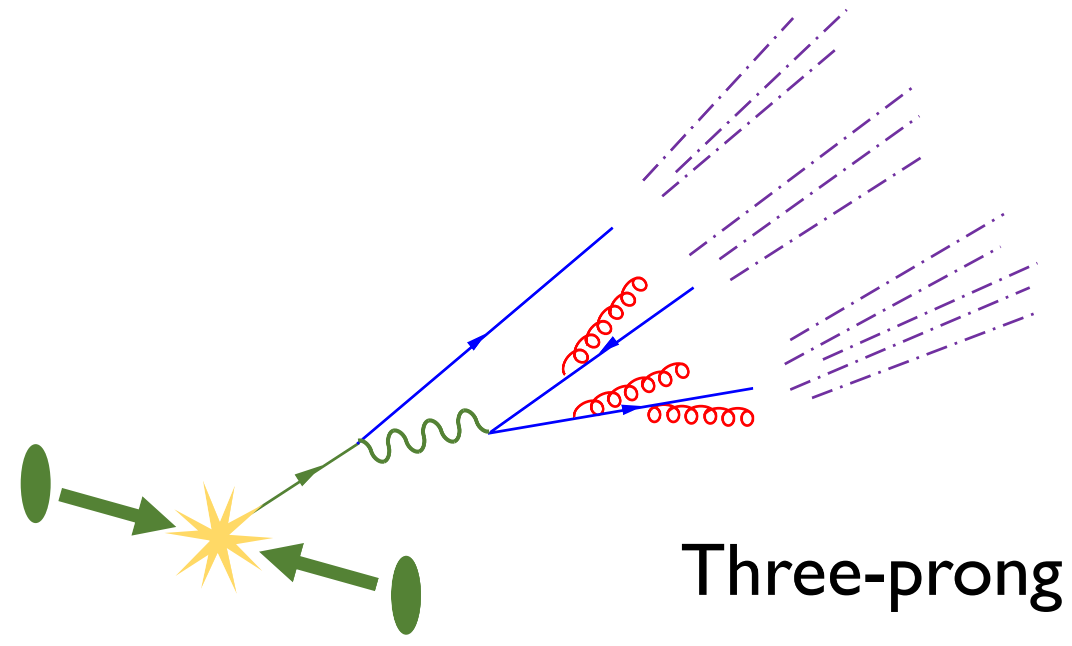
47.1 GeV

hadronization into detector-stable particles

27.0 GeV

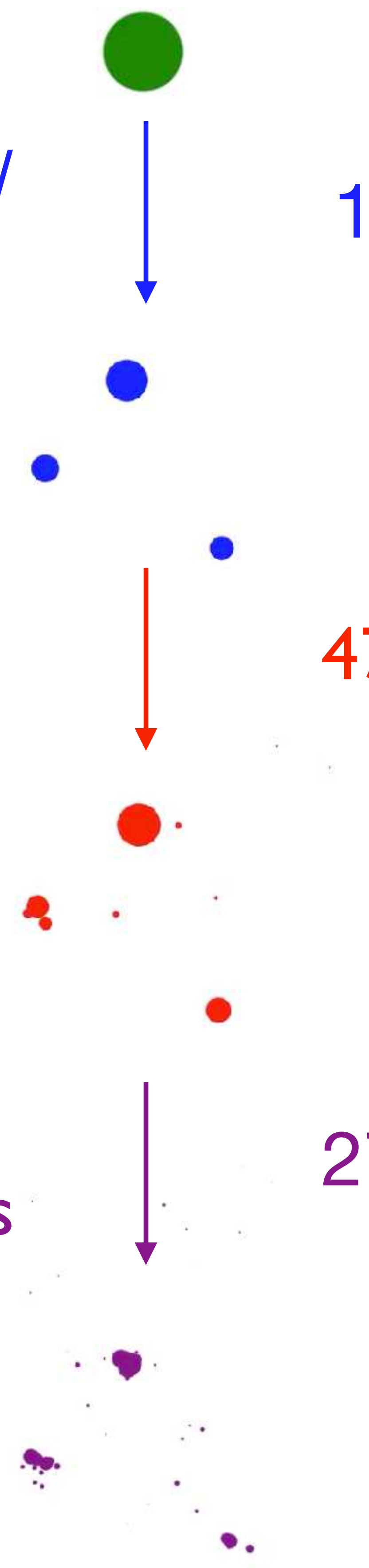


Visualizing Jet Formation – Top Jets



Three-prong decay

500 GeV top quark



decay to W (decays to quark/anti-quark) and a b quark

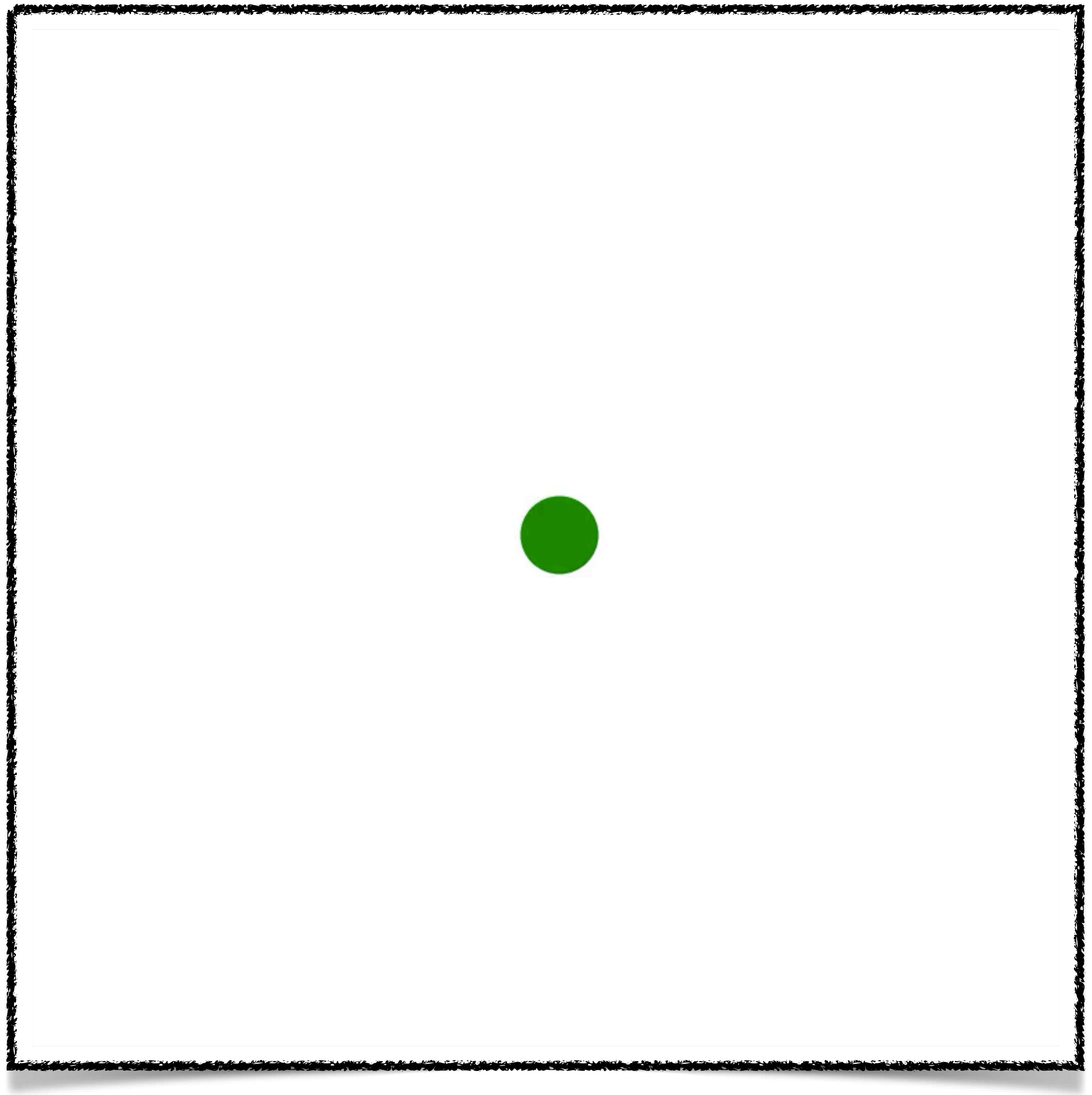
161.1 GeV

fragmentation into quarks and gluons

47.1 GeV

hadronization into detector-stable particles

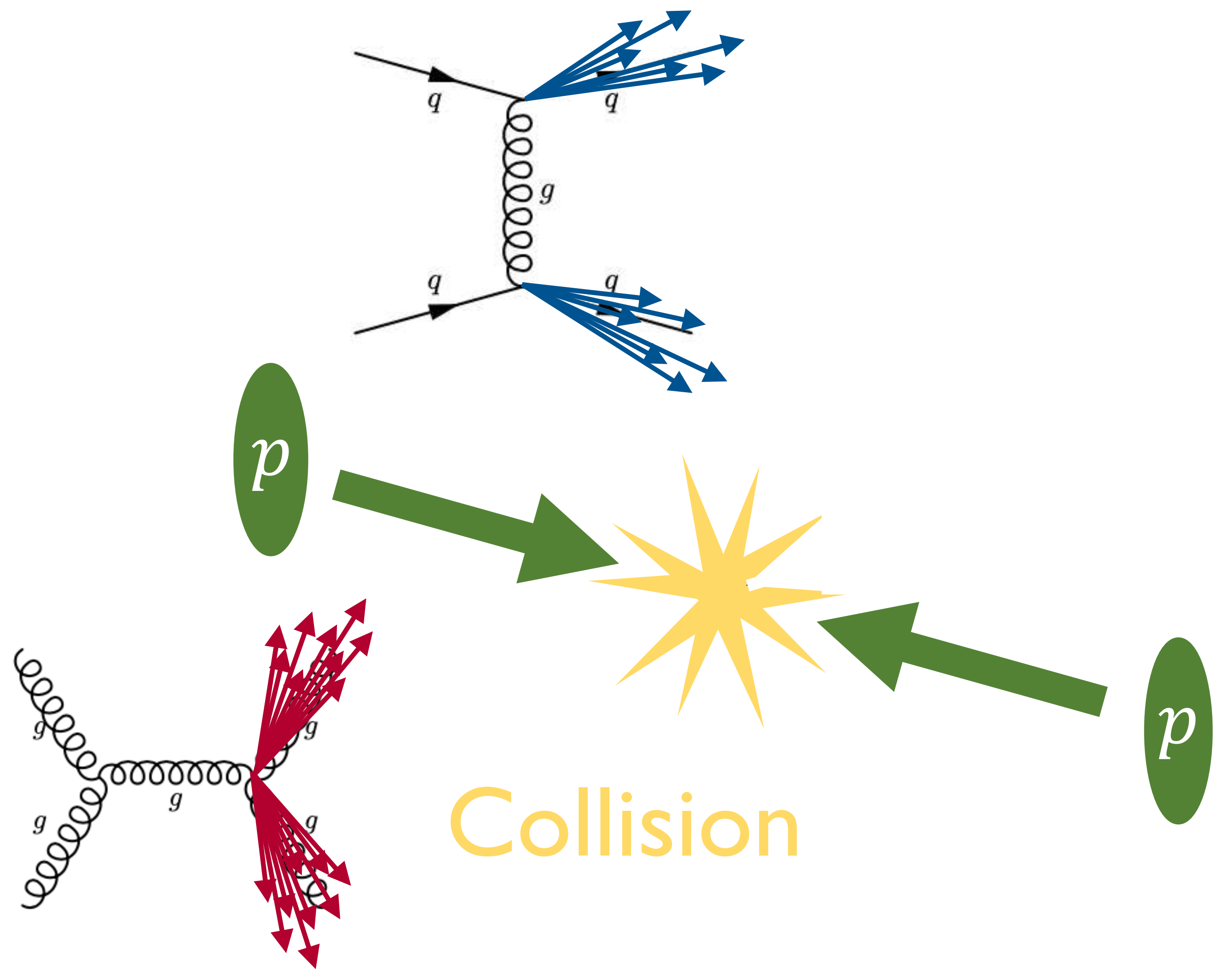
27.0 GeV



Jet Formation in Theory

Hard collision – interesting high-energy dynamics

Perturbative quantum field theory, Feynman diagrams



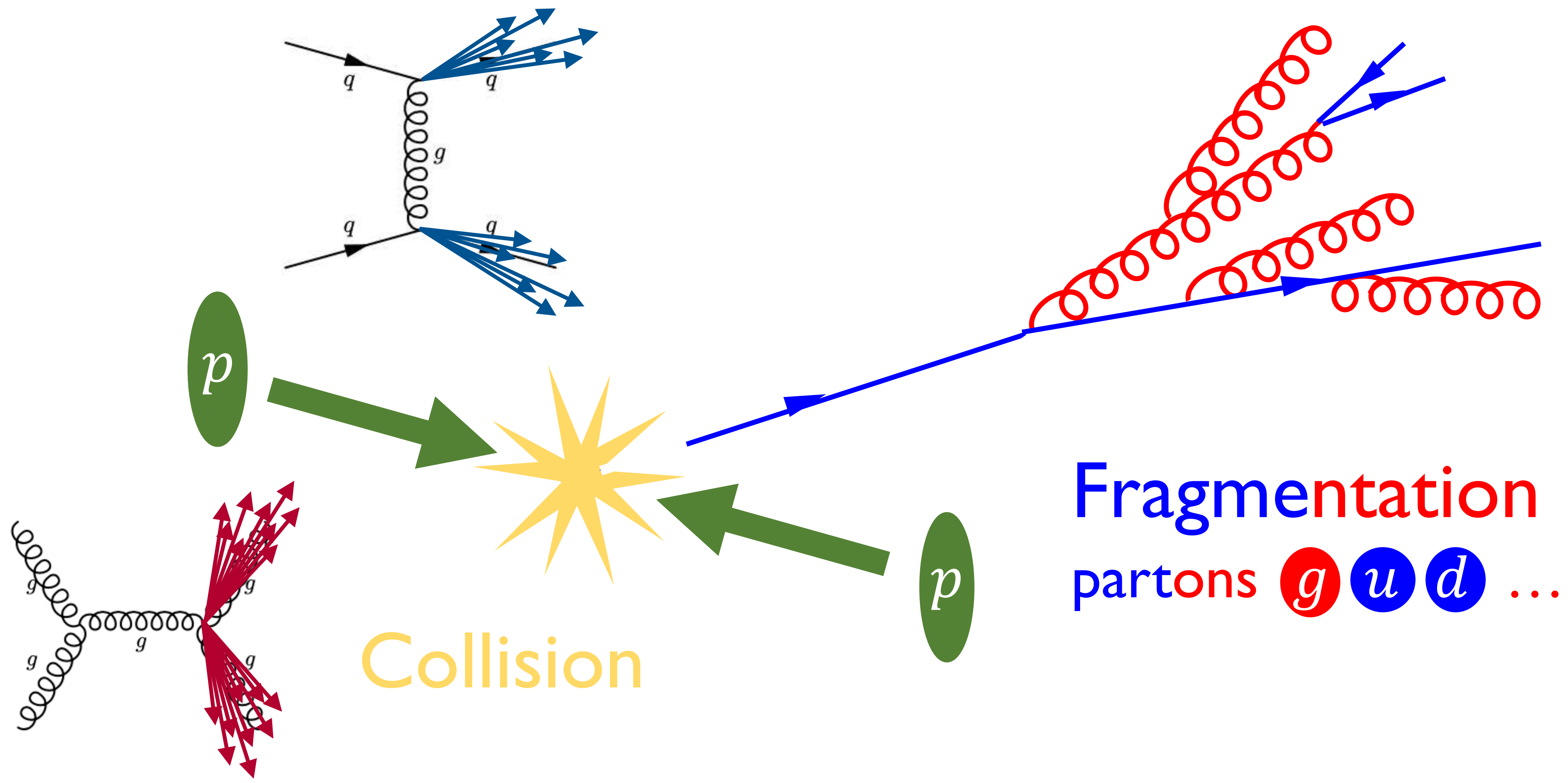
Jet Formation in Theory

Hard collision – interesting high-energy dynamics

Perturbative quantum field theory, Feynman diagrams

Fragmentation – additional gluon radiation

Semi-classical parton shower, effective field theory



Jet Formation in Theory

Hard collision – interesting high-energy dynamics

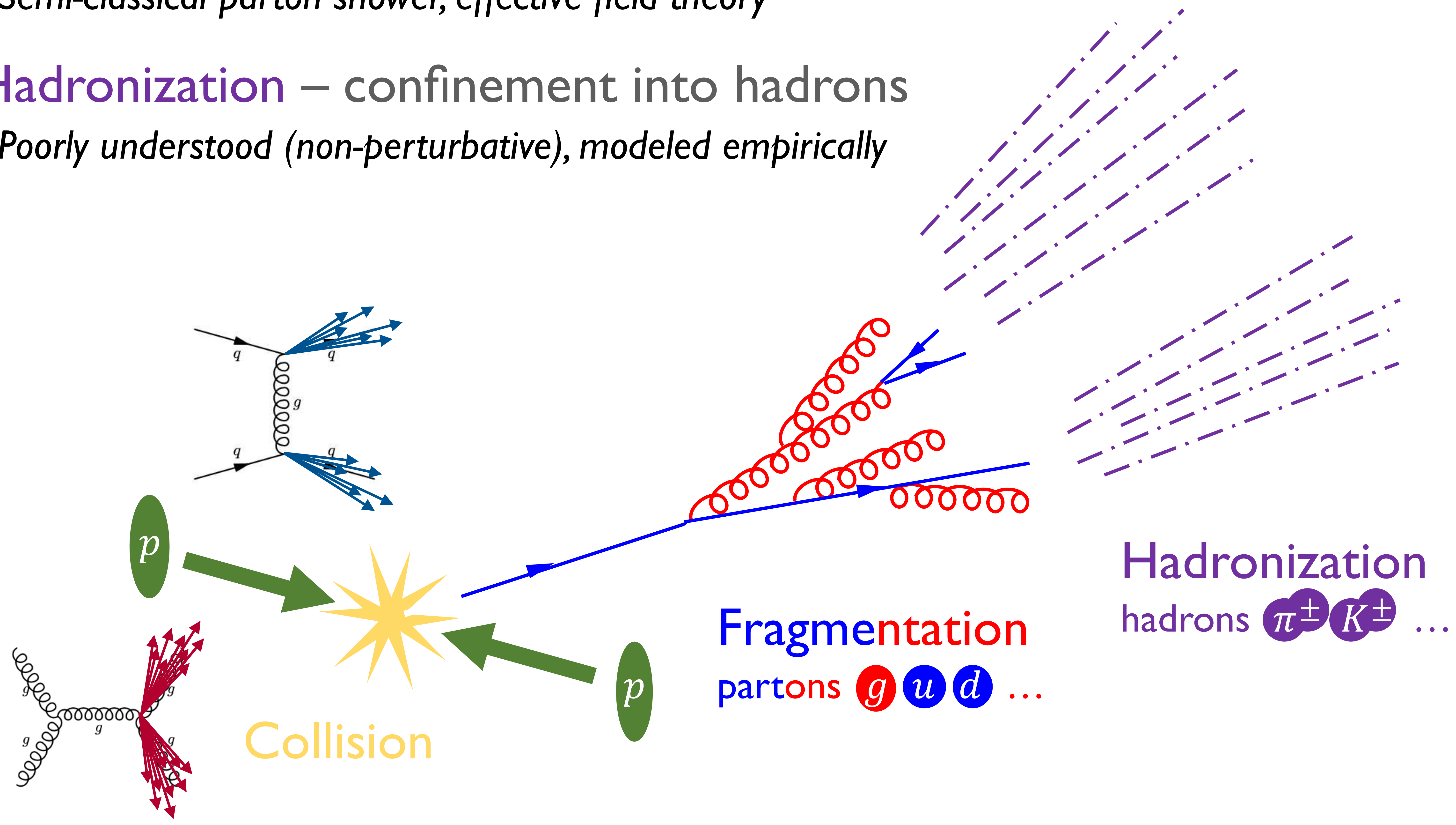
Perturbative quantum field theory, Feynman diagrams

Fragmentation – additional gluon radiation

Semi-classical parton shower, effective field theory

Hadronization – confinement into hadrons

Poorly understood (non-perturbative), modeled empirically



Jet Formation in Theory

Hard collision – interesting high-energy dynamics

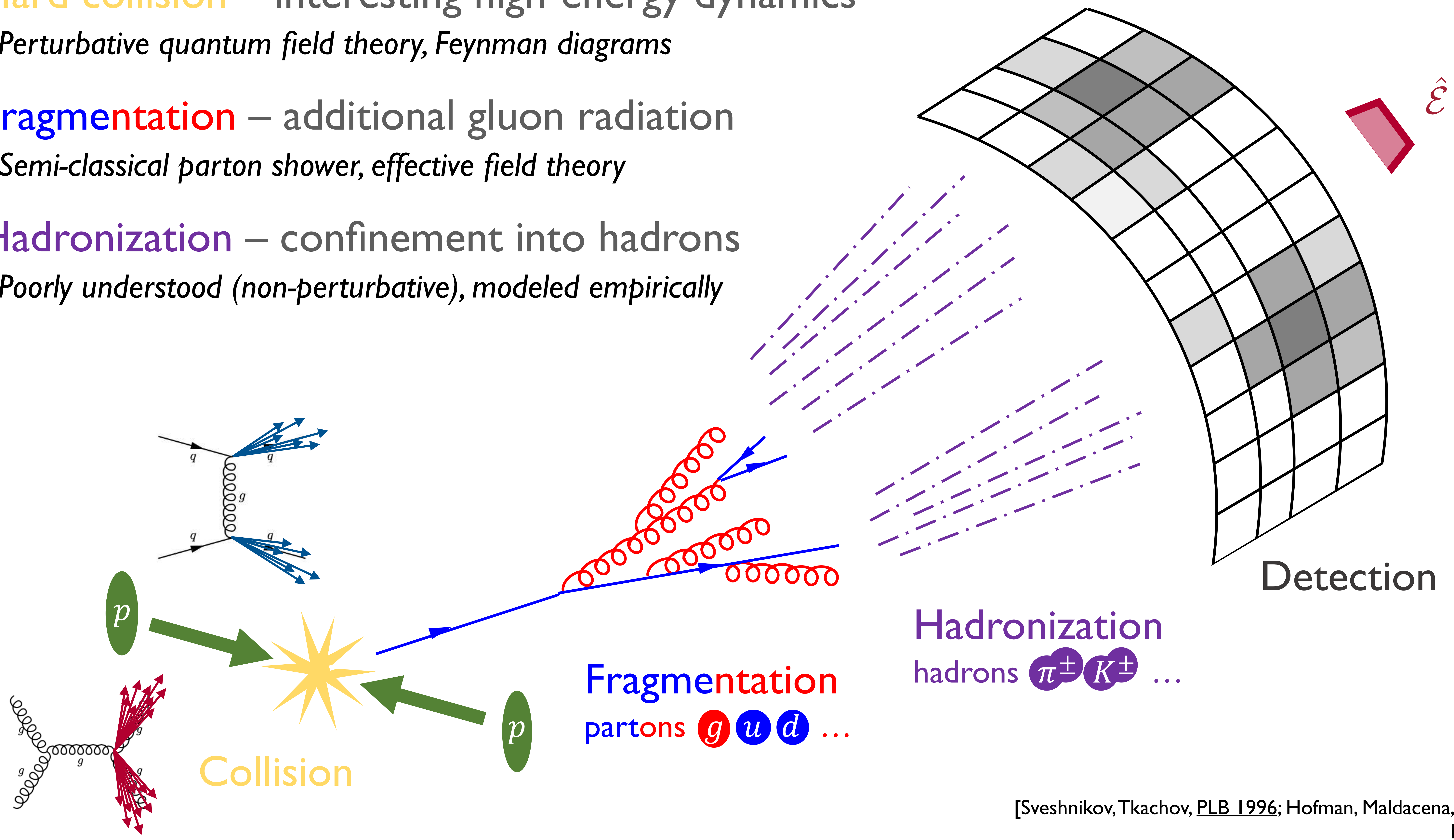
Perturbative quantum field theory, Feynman diagrams

Fragmentation – additional gluon radiation

Semi-classical parton shower, effective field theory

Hadronization – confinement into hadrons

Poorly understood (non-perturbative), modeled empirically



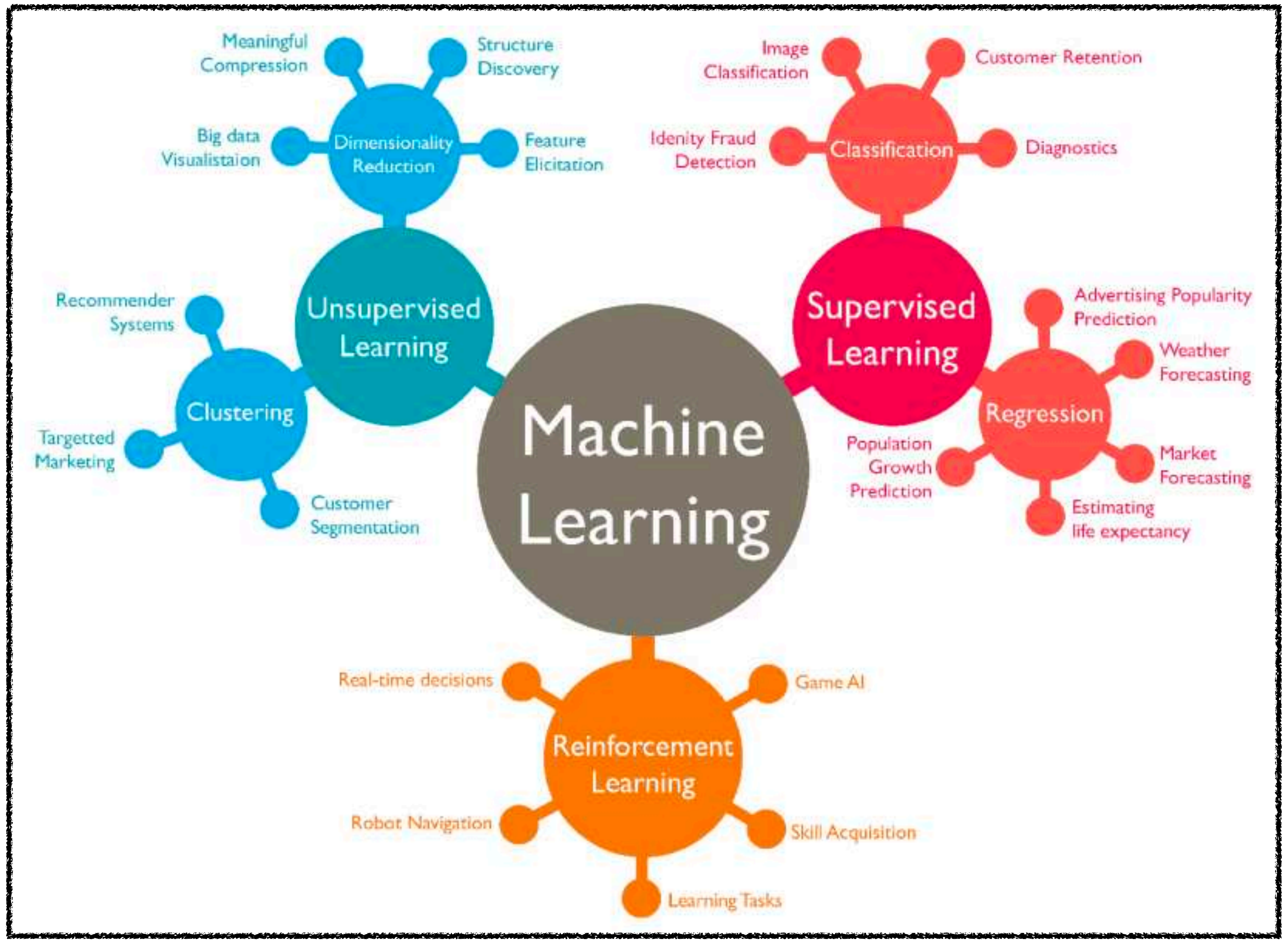
$$\hat{\mathcal{E}} \simeq \lim_{t\to\infty} \hat{n}_i T^{0i}(t, vt\hat{n})$$

Stress-energy flow

Operator that probes direction and amount of outgoing energy

[Sveshnikov, Tkachov, PLB 1996; Hofman, Maldacena, JHEP 2008; Mateu, Stewart, Thaler, PRD 2013; Dixon, PTK, Moult, Thaler, Zhu, *to appear soon*]

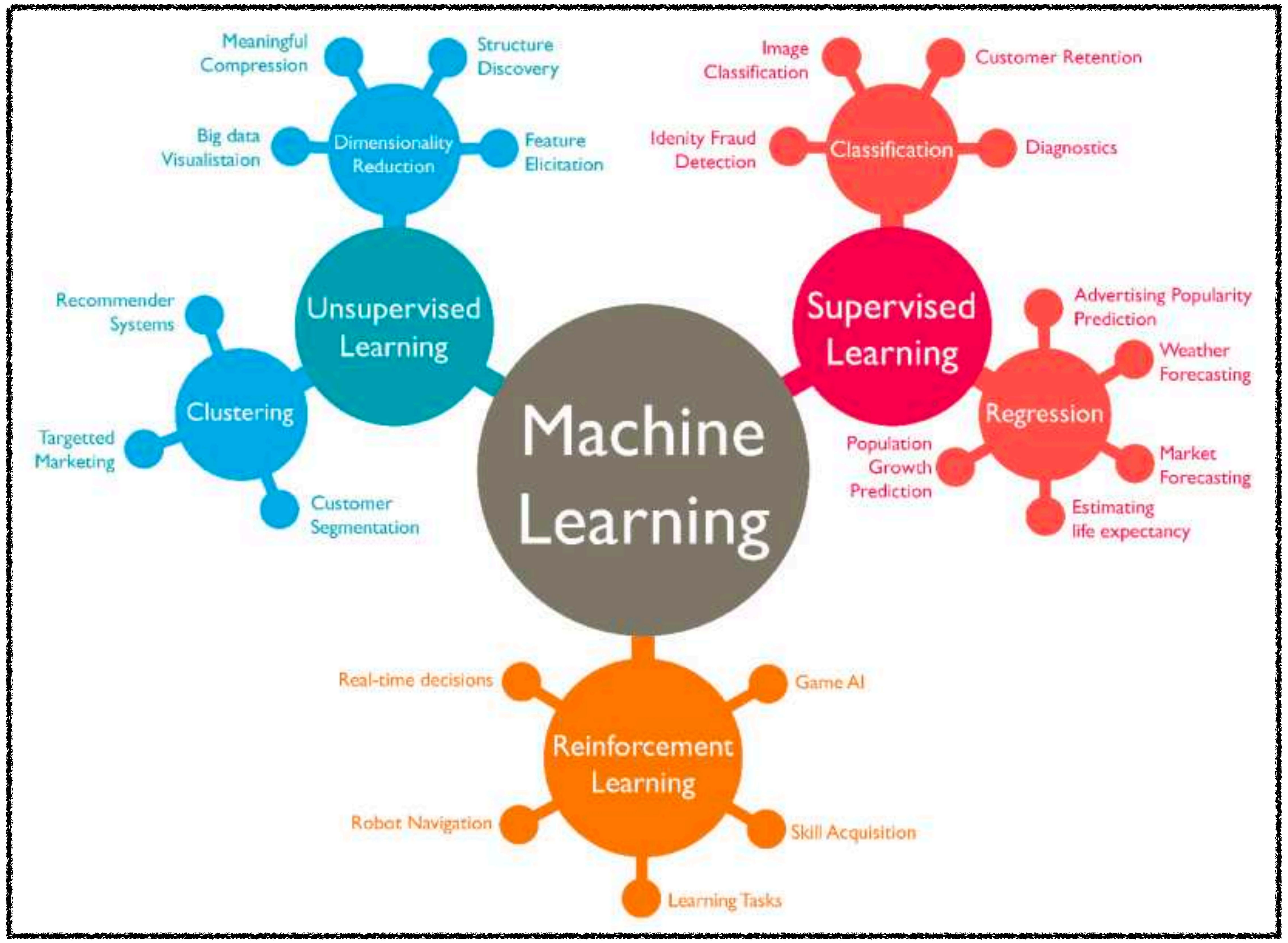
Introduction to Machine Learning



Introduction to Machine Learning

Machine learning comprises statistical algorithms and techniques designed to meaningfully engage with data

$$|\text{machine learning}\rangle \simeq |\text{data science}\rangle = \frac{|\text{statistics}\rangle + |\text{computer science}\rangle}{\sqrt{2}}$$

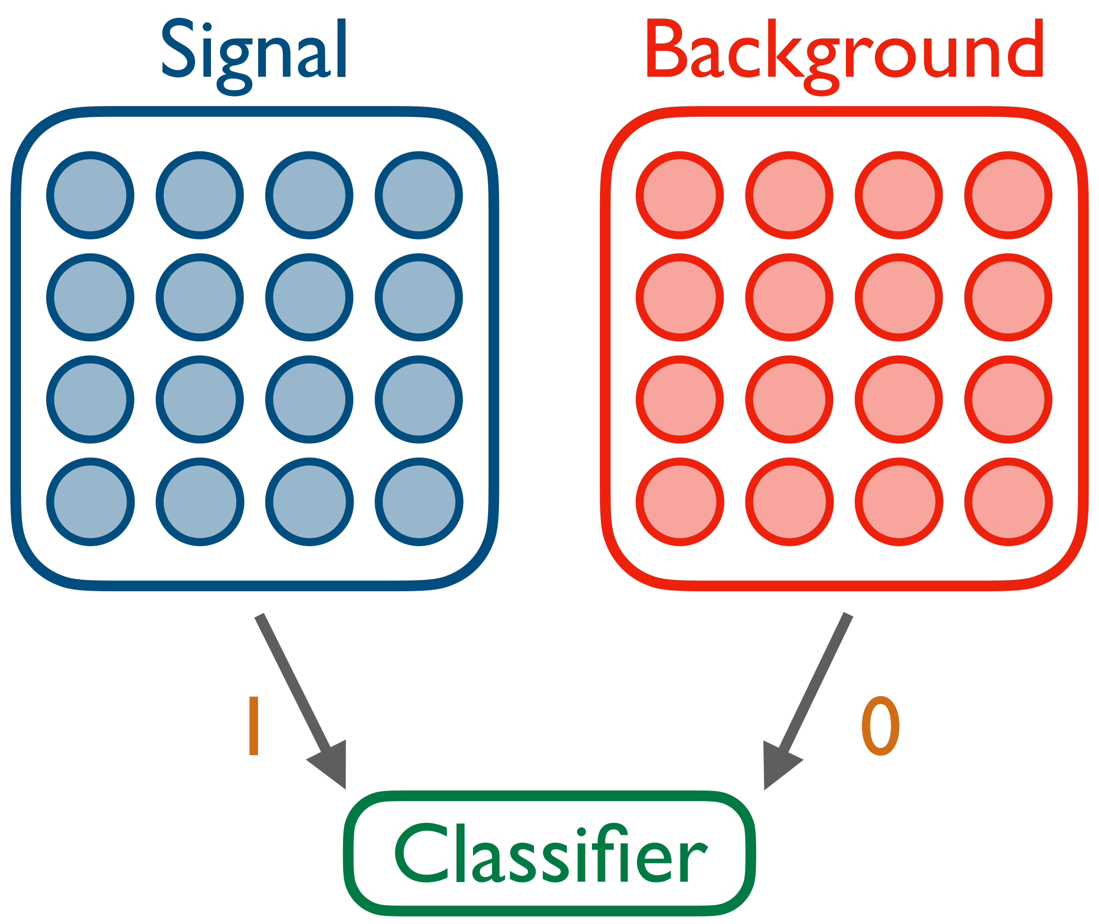


Introduction to Machine Learning

Machine learning comprises statistical algorithms and techniques designed to meaningfully engage with data

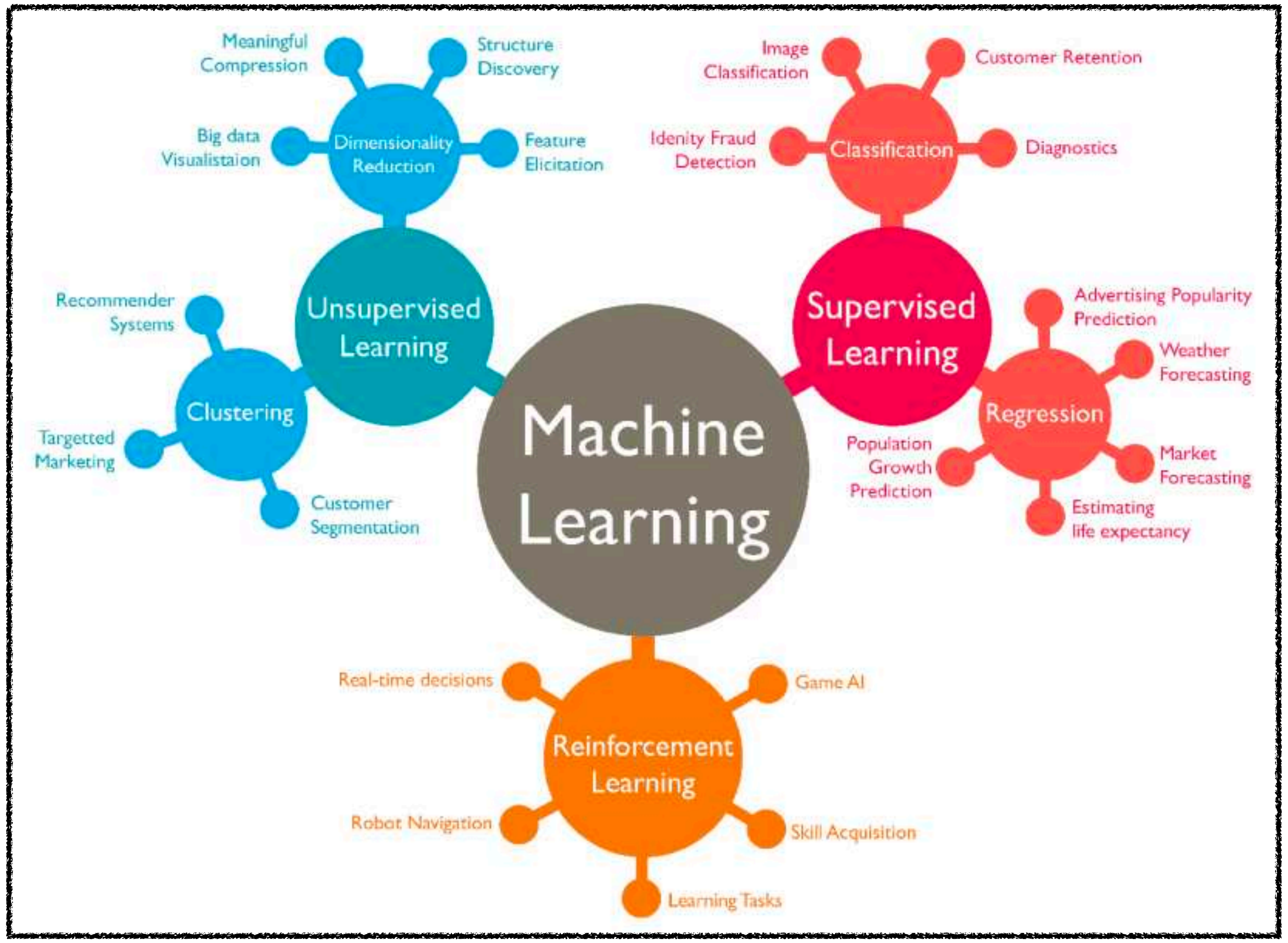
$$|\text{machine learning}\rangle \simeq |\text{data science}\rangle = \frac{|\text{statistics}\rangle + |\text{computer science}\rangle}{\sqrt{2}}$$

Example – Binary Classification



Common paradigm – minimize a loss function

$$\text{loss} = \left\langle (\text{model}_{\vec{w}}(\text{inputs}) - \text{outputs})^2 \right\rangle$$



Machine Learning Considerations

The Power of ML

- *Automatic feature extraction*

  Ensures relevant features are not missed

- *Asymptotically optimizes performance*

  Provides useful/practical statistical power

- *Interpolation in high-dimensional spaces*

  Combats the curse of dimensionality

Machine Learning Considerations

The Power of ML

Comes at a cost

- ***Automatic feature extraction***

  Ensures relevant features are not missed

  Cannot easily convey what features are used

- ***Asymptotically optimizes performance***

  Provides useful/practical statistical power

  Training is difficult with few global guarantees

- ***Interpolation in high-dimensional spaces***

  Combats the curse of dimensionality

  Loses analytic understandability/tractability

Machine Learning Considerations – in Particle Physics

The Power of ML

Comes at a cost

- *Automatic feature extraction*

  Ensures relevant features are not missed

  Cannot easily convey what features are used

- *Asymptotically optimizes performance*

  Provides useful/practical statistical power

  Training is difficult with few global guarantees

- *Interpolation in high-dimensional spaces*

  Combats the curse of dimensionality

  Loses analytic understandability/tractability

⇨

- *Available data*

  Data source, number of samples, labels, reliability

- *Learning paradigm*

  Fully/weakly/un-supervised, classification/regression/generation

- *Inputs and outputs*

  Size/shape, symmetries, dimensionality

- *Model architecture*

  Expressibility, loss function, hyperparameters, validation/testing

- *Deployment strategy*

  Model implementation, training/evaluation speed, uncertainties

Machine Learning Considerations – in Particle Physics

The Power of ML

Comes at a cost

- *Automatic feature extraction*

  Ensures relevant features are not missed

  Cannot easily convey what features are used

- *Asymptotically optimizes performance*

  Provides useful/practical statistical power

  Training is difficult with few global guarantees

- *Interpolation in high-dimensional spaces*

  Combats the curse of dimensionality

  Loses analytic understandability/tractability

⇨

- *Available data*

  Data source, number of samples, labels, reliability

- *Learning paradigm*

  Fully/weakly/un-supervised, classification/regression/generation

- *Inputs and outputs*

  Size/shape, symmetries, dimensionality

- *Model architecture*

  Expressibility, loss function, hyperparameters, validation/testing

- *Deployment strategy*

  Model implementation, training/evaluation speed, uncertainties

Machine Learning Considerations – in Particle Physics

The Power of ML

Comes at a cost

- ***Automatic feature extraction***

  Ensures relevant features are not missed

  Cannot easily convey what features are used

- ***Asymptotically optimizes performance***

  Provides useful/practical statistical power

  Training is difficult with few global guarantees

- ***Interpolation in high-dimensional spaces***

  Combats the curse of dimensionality

  Loses analytic understandability/tractability

- ***Available data***

  Data source, number of samples, labels, reliability

- ***Learning paradigm***

  Fully/weakly/un-supervised, classification/regression/generation

- ***Inputs and outputs***

  Size/shape, symmetries, dimensionality

- ***Model architecture***

  Expressibility, loss function, hyperparameters, validation/testing

- ***Deployment strategy***

  Model implementation, training/evaluation speed, uncertainties

Machine Learning Considerations – in Particle Physics

The Power of ML

Comes at a cost

- *Automatic feature extraction*

  Ensures relevant features are not missed

  Cannot easily convey what features are used

- *Asymptotically optimizes performance*

  Provides useful/practical statistical power

  Training is difficult with few global guarantees

- *Interpolation in high-dimensional spaces*

  Combats the curse of dimensionality

  Loses analytic understandability/tractability

⇨

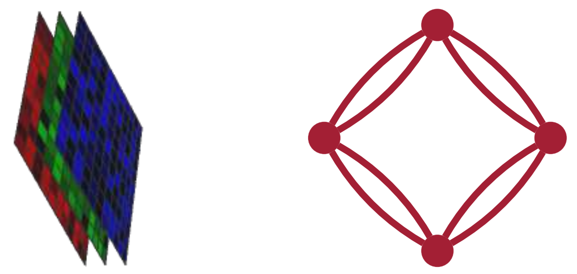
- *Available data*

  Data source, number of samples, labels, reliability

- *Learning paradigm*

  Fully/weakly/un-supervised, classification/regression/generation

- *Inputs and outputs*



  Size/shape, symmetries, dimensionality

- *Model architecture*

  Expressibility, loss function, hyperparameters, validation/testing

- *Deployment strategy*

  Model implementation, training/evaluation speed, uncertainties

Machine Learning Considerations – in Particle Physics

The Power of ML

Comes at a cost

- *Automatic feature extraction*

  Ensures relevant features are not missed

  Cannot easily convey what features are used

- *Asymptotically optimizes performance*

  Provides useful/practical statistical power

  Training is difficult with few global guarantees

- *Interpolation in high-dimensional spaces*

  Combats the curse of dimensionality

  Loses analytic understandability/tractability

⇨

- *Available data*

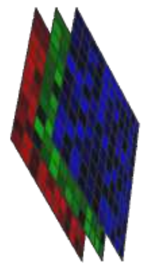
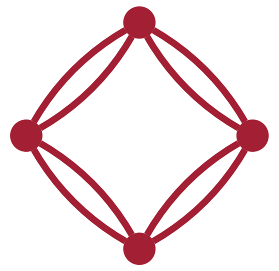
  Data source, number of samples, labels, reliability

- *Learning paradigm*



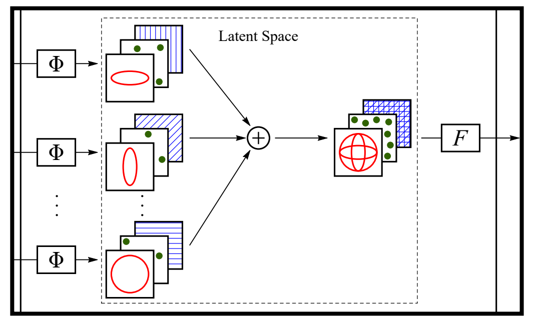
  Fully/weakly/un-supervised, classification/regression/generation

- *Inputs and outputs*

  Size/shape, symmetries, dimensionality

- *Model architecture*



  Expressibility, loss function, hyperparameters, validation/testing

- *Deployment strategy*

  Model implementation, training/evaluation speed, uncertainties

Machine Learning Considerations – in Particle Physics

The Power of ML

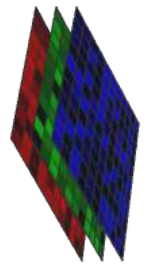
Comes at a cost

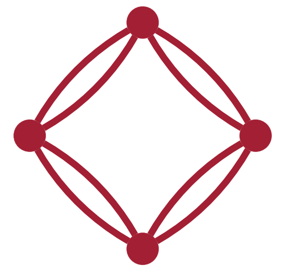
- *Automatic feature extraction*
  Ensures relevant features are not missed
  Cannot easily convey what features are used
- *Asymptotically optimizes performance*
  Provides useful/practical statistical power
  Training is difficult with few global guarantees
- *Interpolation in high-dimensional spaces*
  Combats the curse of dimensionality
  Loses analytic understandability/tractability

⇨

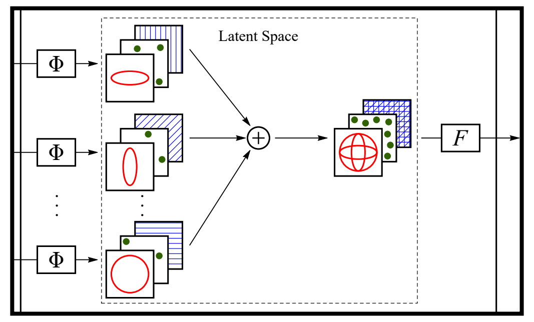
- *Available data*
  Data source, number of samples, labels, reliability
- *Learning paradigm*

  Fully/weakly/un-supervised, classification/regression/generation
- *Inputs and outputs*



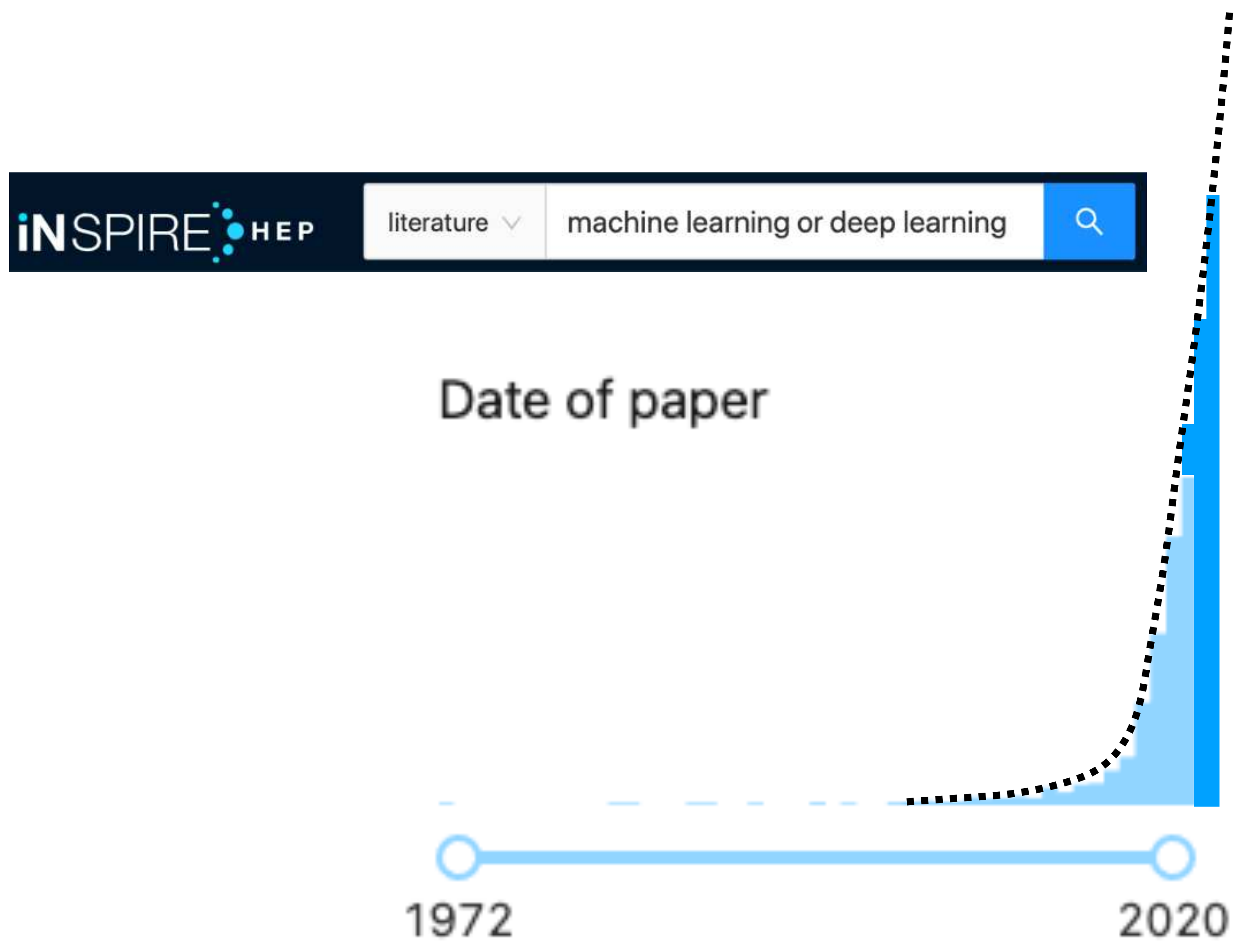


  Size/shape, symmetries, dimensionality
- *Model architecture*



  Expressibility, loss function, hyperparameters, validation/testing
- *Deployment strategy*
  Model implementation, training/evaluation speed, uncertainties

Thoughts on Machine Learning



- *Machine learning will be essential in maximizing HEP potential*
  - We should capitalize on the opportunity to optimize
  - ML is both a computational tool and a useful formalism/language

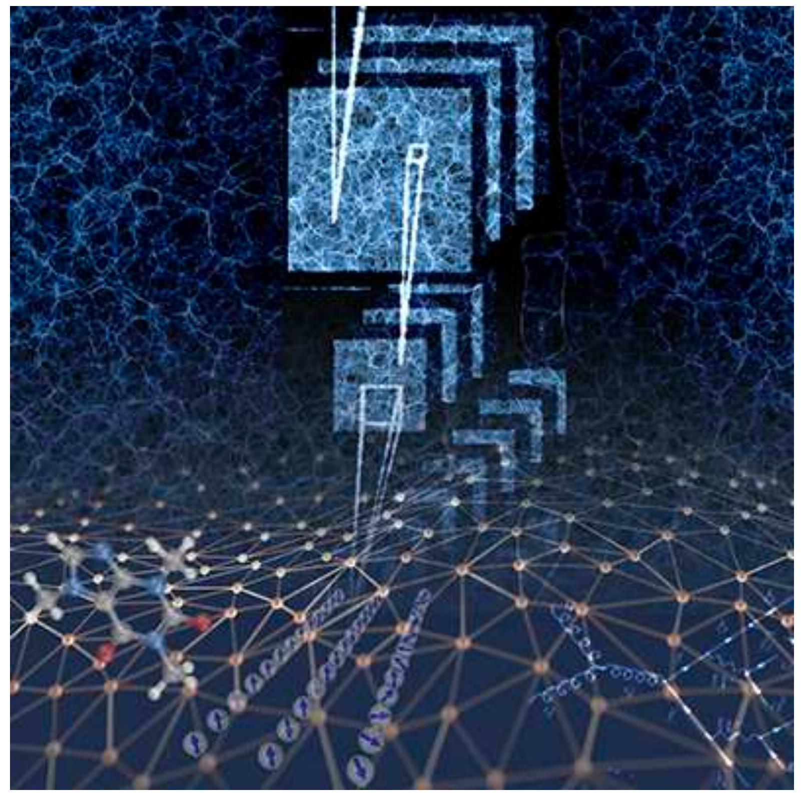
HEPML-LivingReview has a thorough and organized list of papers

- *High-energy physics can benefit ML*
  - e.g. EFNs are weighted deep sets, EFN2/PFN2 will have broader applications
  - NSF Institute for Artificial Intelligence and Fundamental Interactions

$$\mathrm{EFN}^{(\ell_1,\ell_2,\ldots)}(\{p_i^\mu\}) = F\left(\underbrace{\sum_i z_i\Phi^{(\ell_1)}(\hat{p}_i)}_{\Phi_1}, \underbrace{\sum_i\sum_j z_iz_j\Phi^{(\ell_2)}(\hat{p}_i,\hat{p}_j)}_{\Phi_2}, \ldots\right)$$

- *Collaboration across traditional lines will enable success*
  - Theorists, experimentalists, and ML experts all can and should contribute

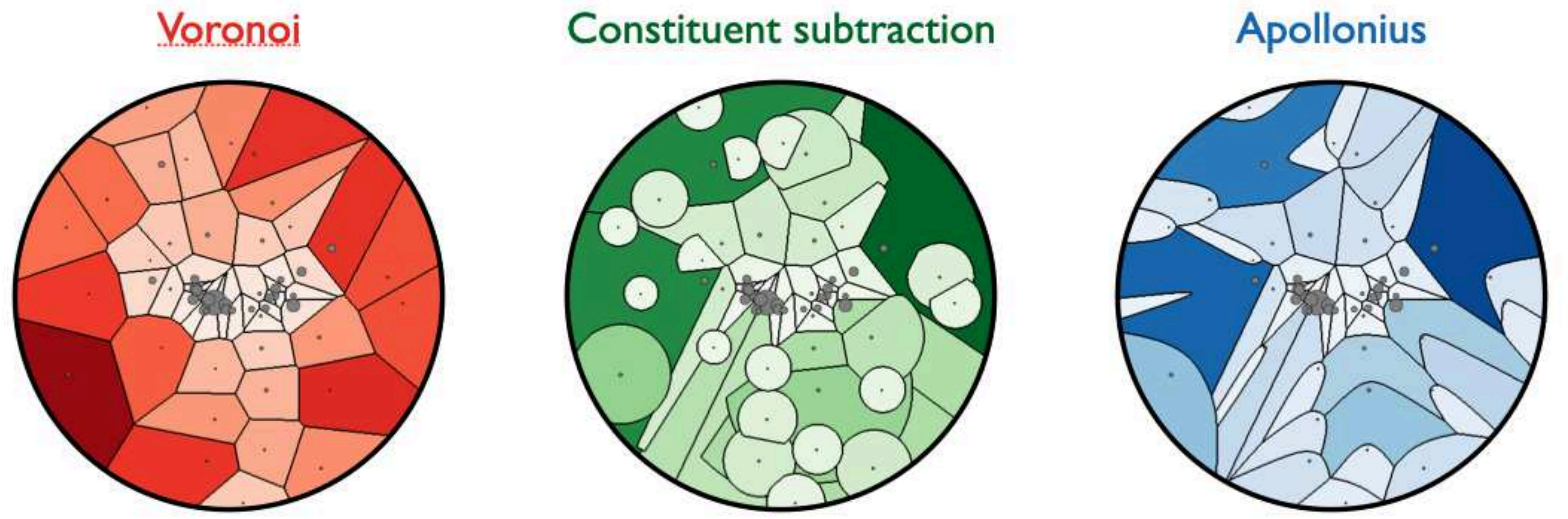
- *Computational best practices can be shared among fields*
  - Software workflows, reproducible analyses, public datasets are critical



[Reviews of Modern Physics Cover December 2019 from Machine learning and the physical sciences]

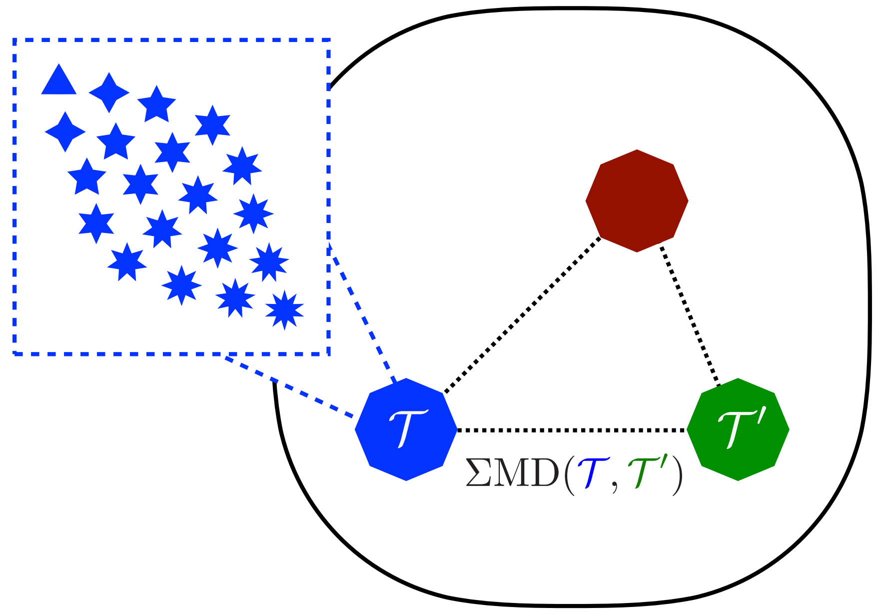
Future Development of Architectures, Algorithms, and Techniques

- *Beyond single-particle point cloud architectures*
  - Pairwise information known to be physically meaningful
  - "EFN2/PFN2" has greater expressivity for tagging and unfolding
- *EMD-inspired techniques for theory and experiment*
  - New and better grooming and pileup mitigation techniques for LHC and beyond
  - Bootstrap "event space" to "theory space"
- *Public datasets provide rich context for testing new methods*
  - Novel search strategies can be tried directly (e.g. dimuon resonances with pT cut)
  - Data preservation critical for maximizing scientific benefit (e.g. ALEPH e^+e^- data)
- *Opportunities pursue novel investigation strategies in HEP*
  - Weighted cross sections probe physics differently than traditional observables
  - e.g. energy-energy correlators, which utilize CFT techniques for QCD calculations



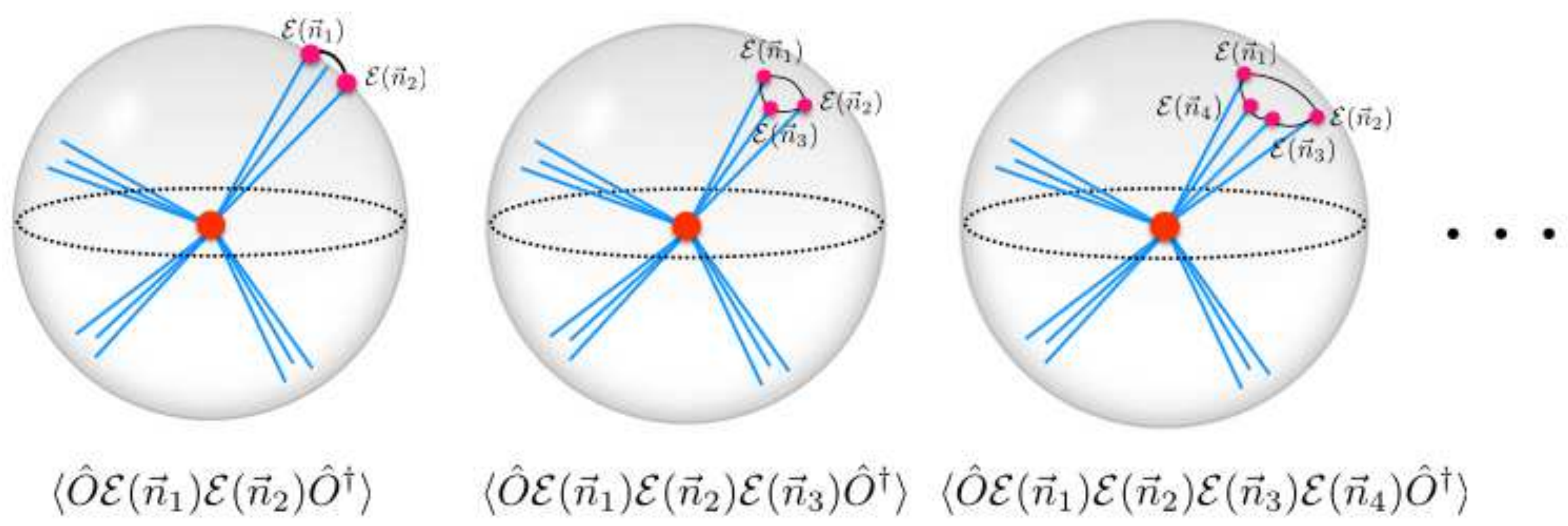
Pileup subtraction methods on example jet from CMS Open Data

(More in backup)



Cartoon of "theory space"

(More in backup)

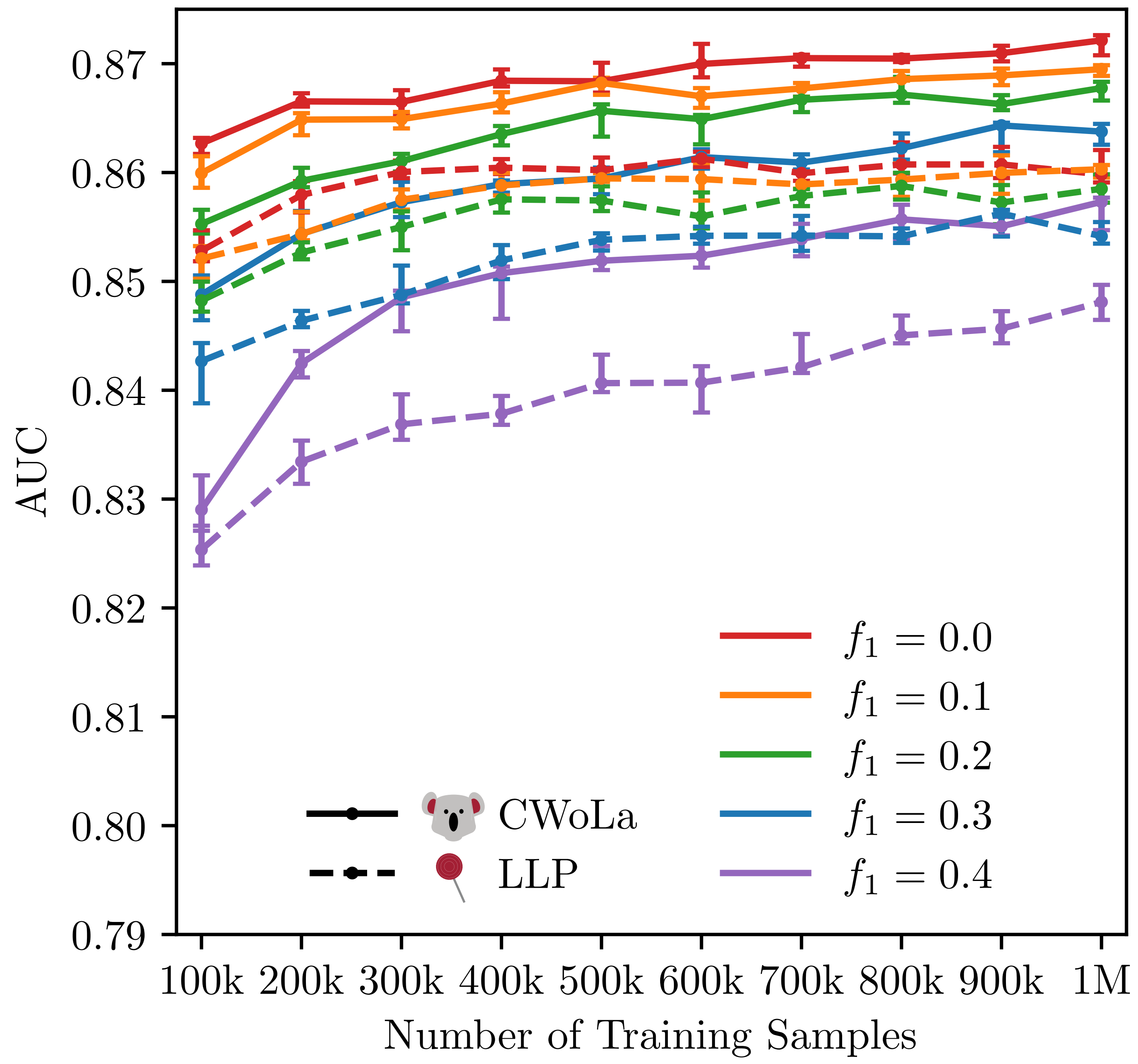


Correlations of energy flow operators on celestial sphere

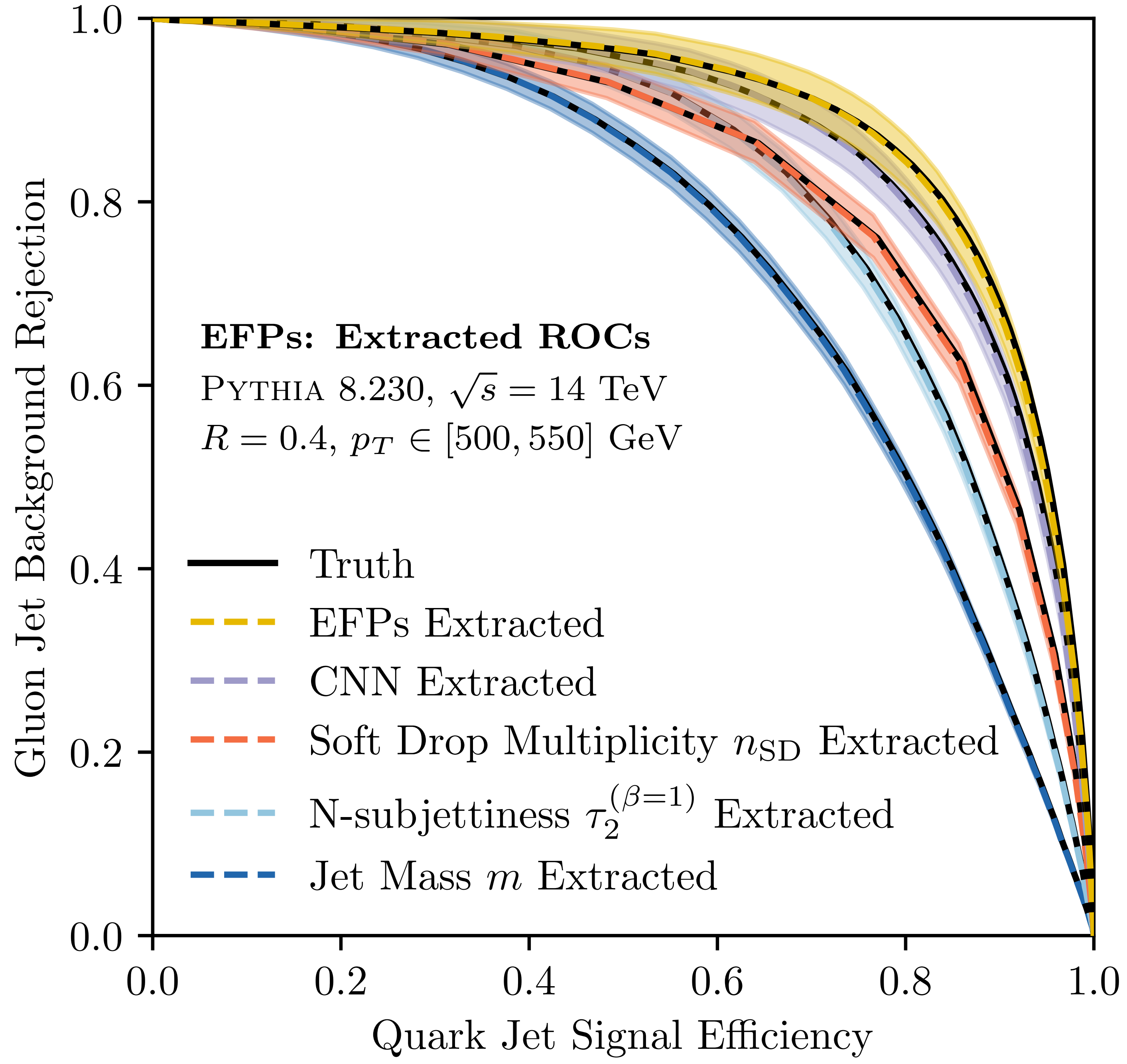
(More in backup)

Training On Data

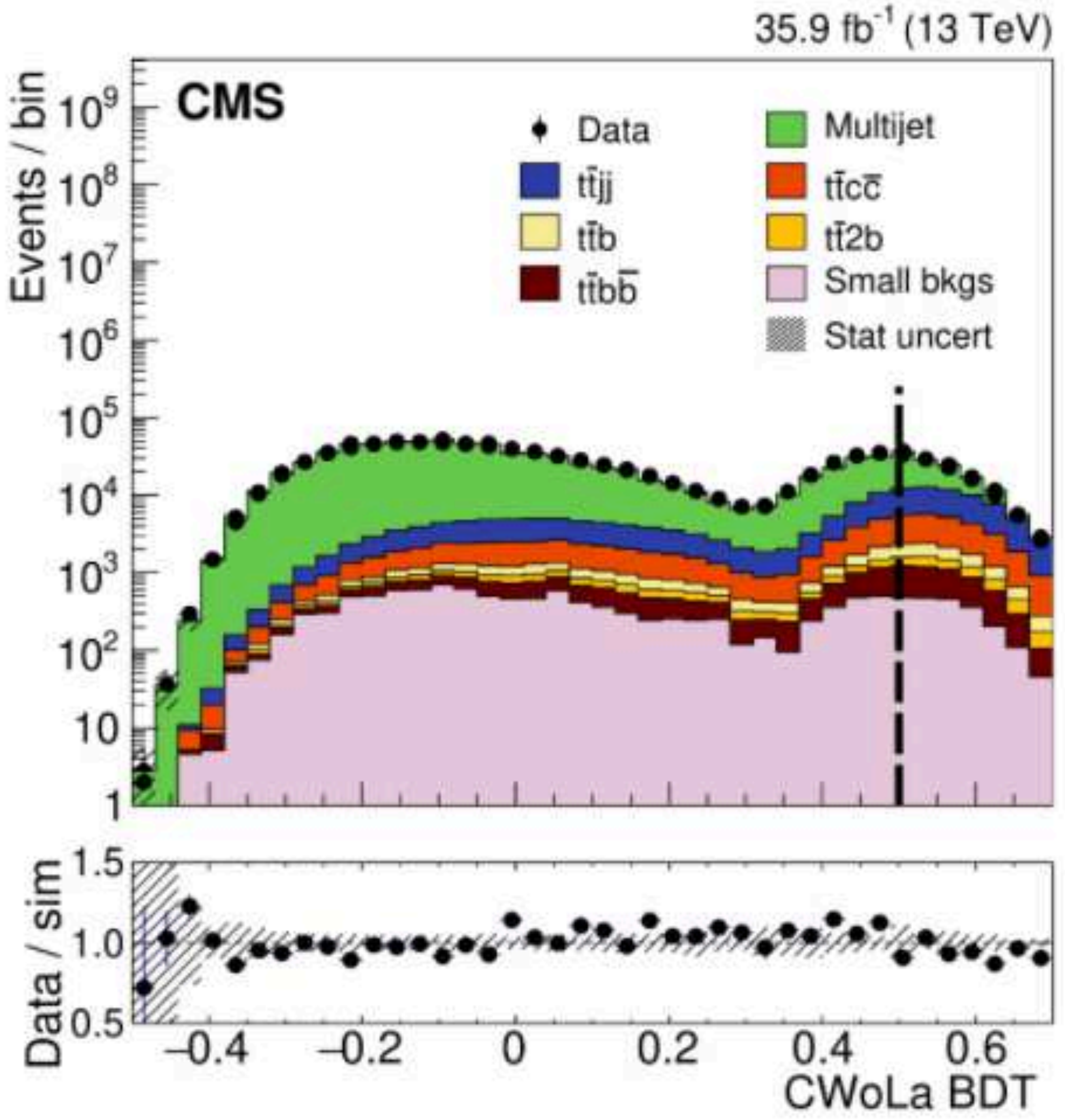
[Metodiev, Nachman, Thaler, JHEP 2017;
PTK, Metodiev, Nachman, Schwartz, PRD 2018;
Metodiev, Thaler, PRL 2018;
PTK, Metodiev, Thaler, JHEP 2018]



CWoLa can be used to train high-dimensional classifiers on mixed samples without labels



Classifiers can be calibrated with topic modeling

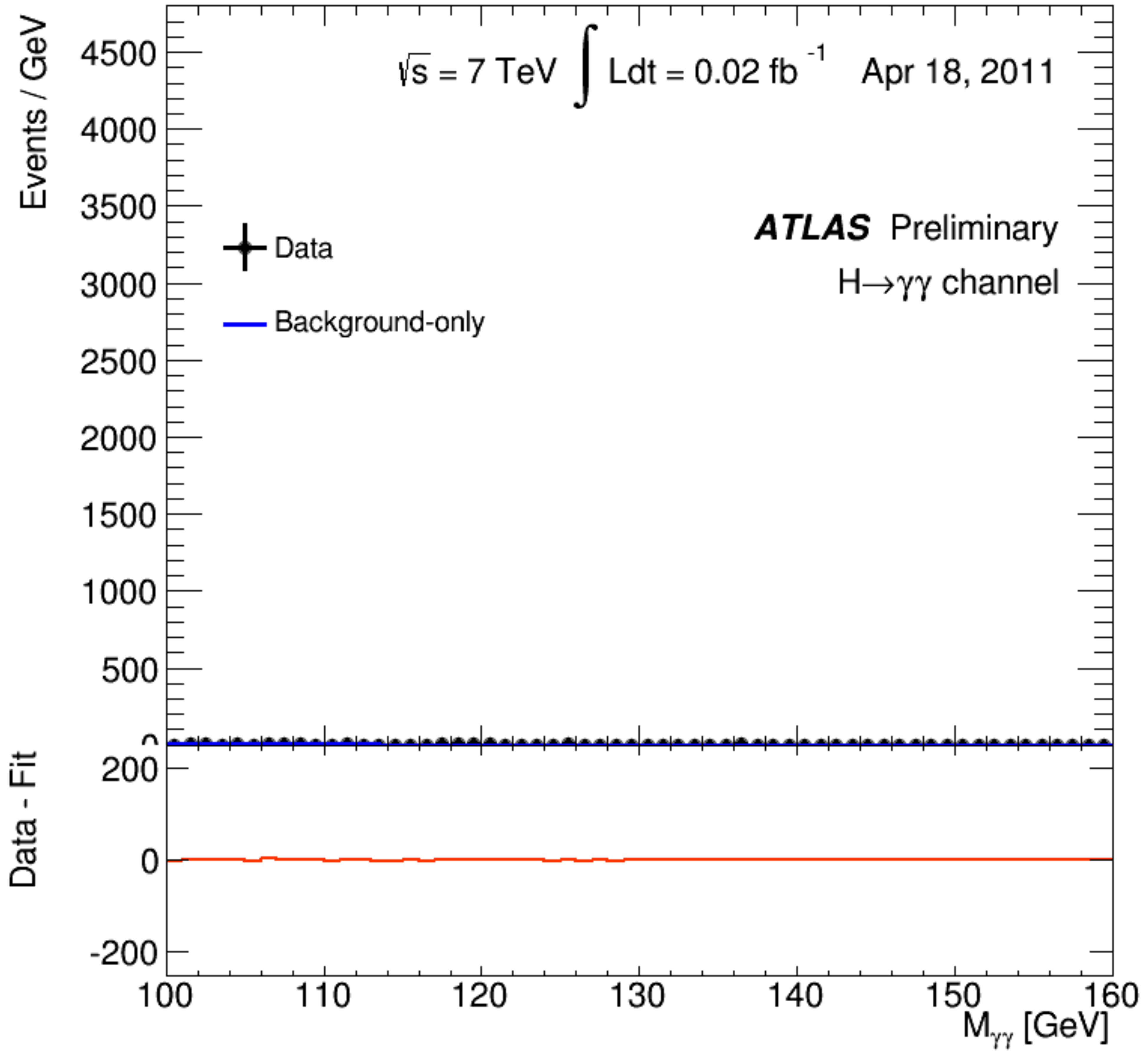


CWoLa used to train BDT in CMS data

[CMS, PLB 2020]

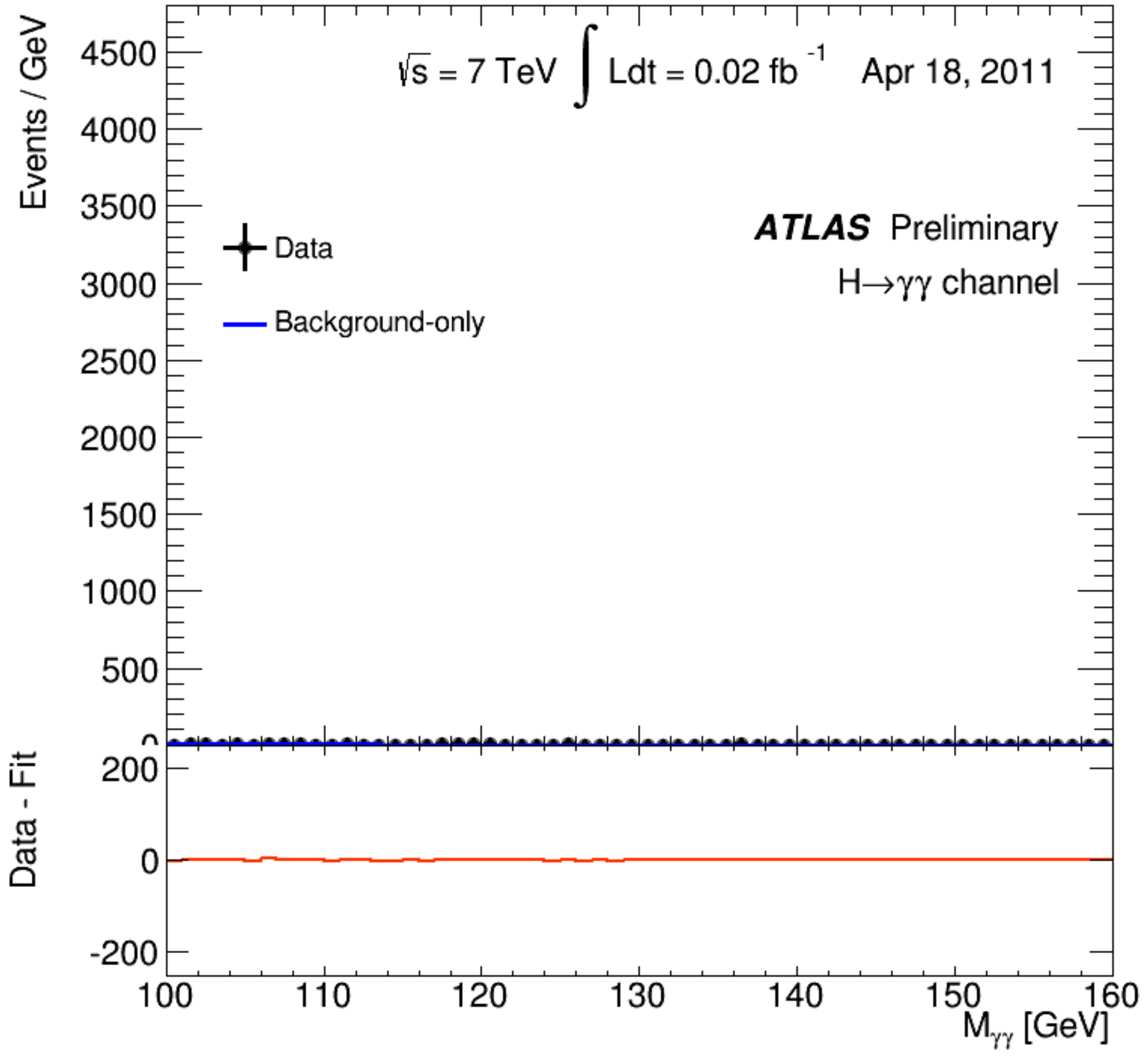
Statistical Ensembles of Events

Histograms show the distribution of events according to some particular feature



Statistical Ensembles of Events

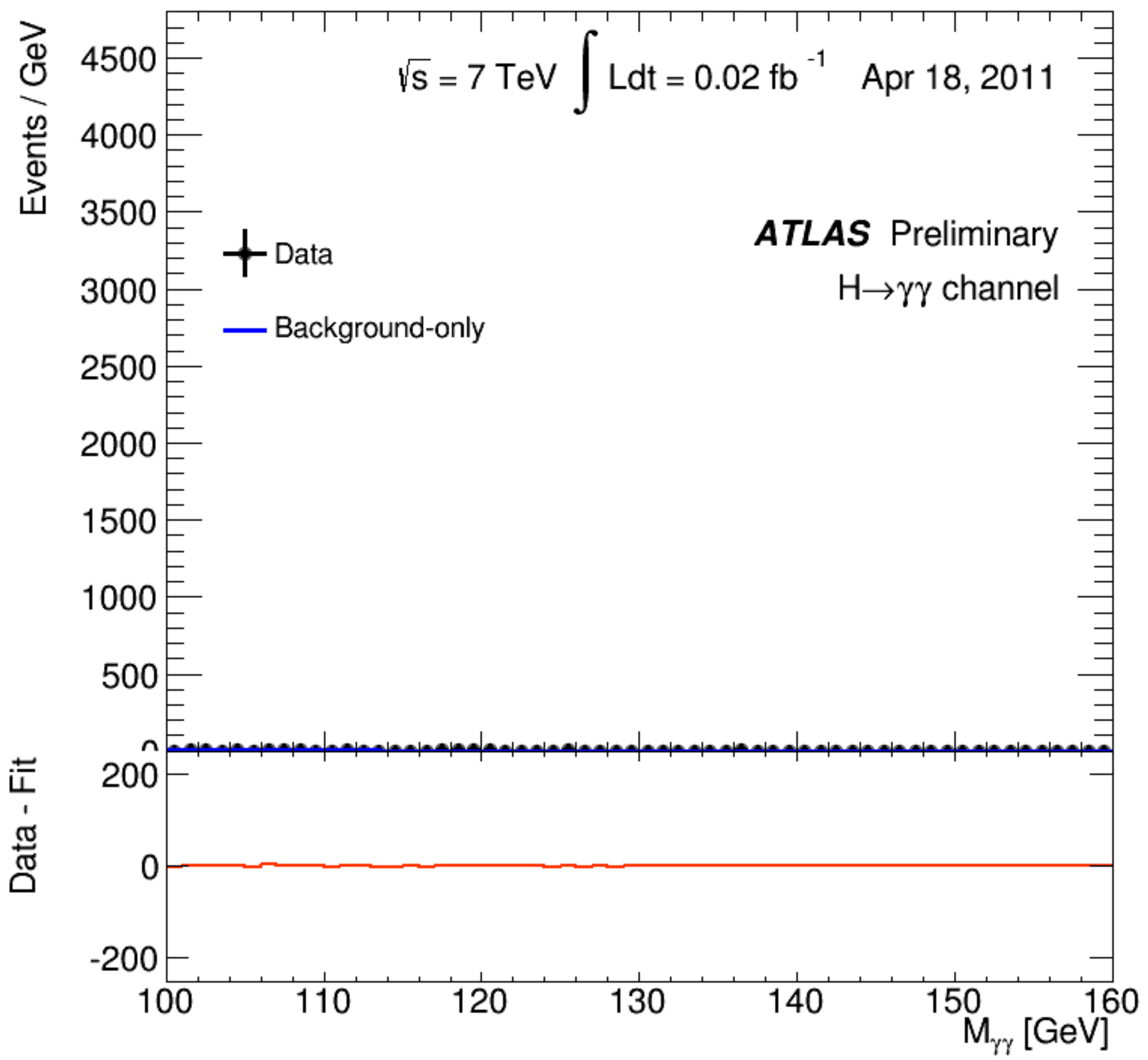
Histograms show the distribution of events according to some particular feature



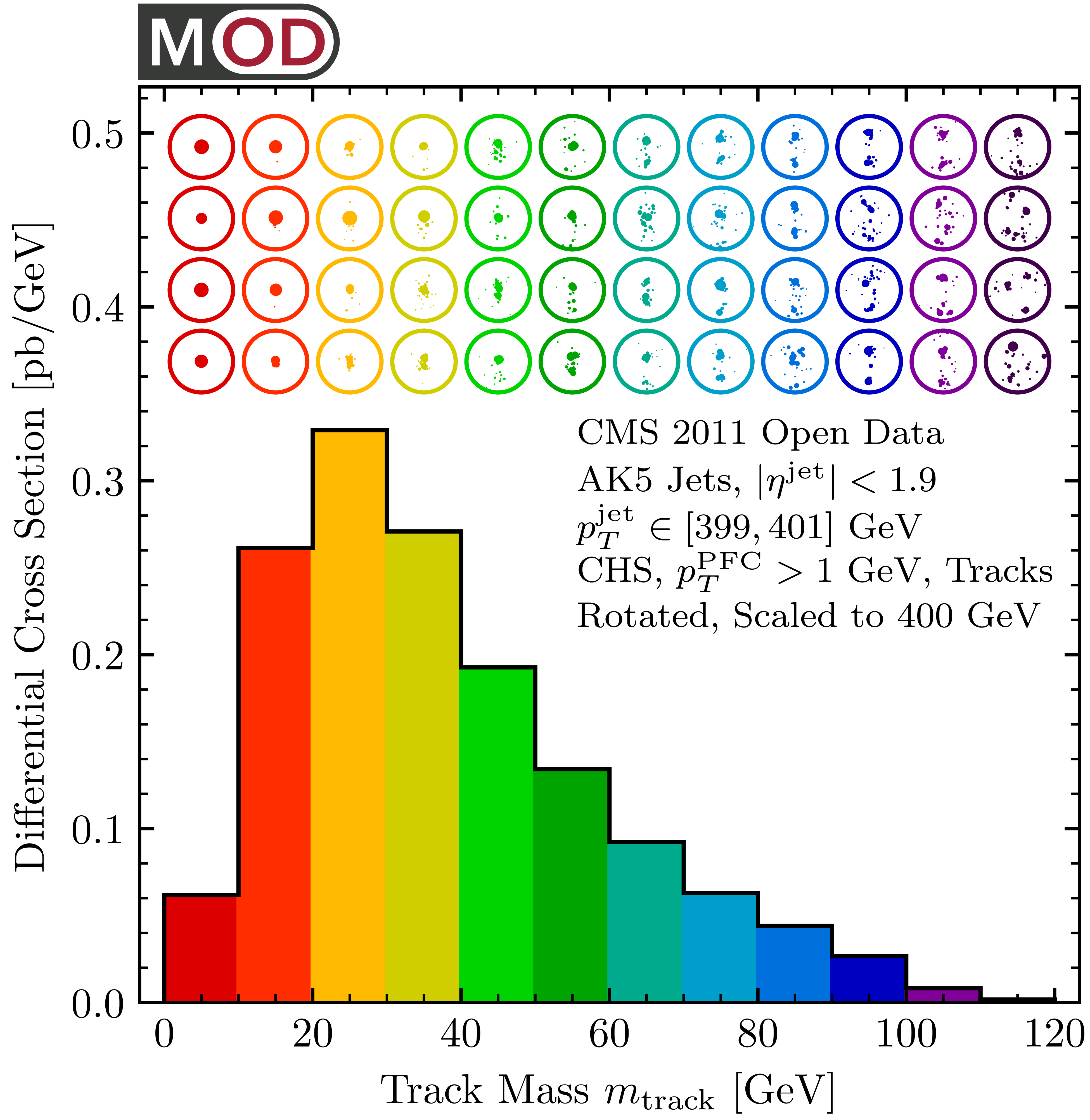
Discovering the Higgs boson from two photons

Statistical Ensembles of Events

Histograms show the distribution of events according to some particular feature



Discovering the Higgs boson from two photons



Showing events by the combined "mass" of their charged particles

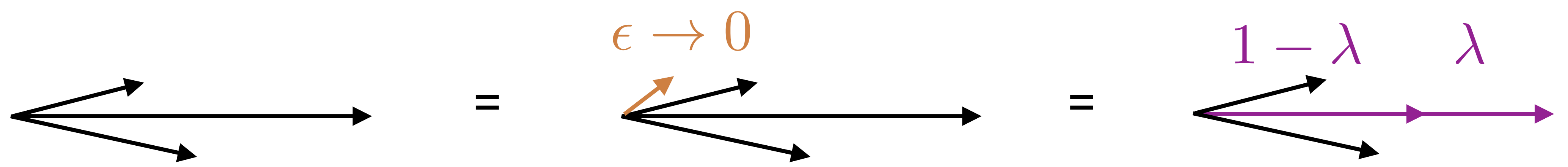
Infrared and Collinear (IRC) Safety

QCD has soft and collinear divergences associated with gluon radiation

$$dP_{i\to ig} \simeq \frac{2\alpha_s}{\pi} C_a \frac{d\theta}{\theta}\frac{dz}{z} \qquad C_q = C_F = 4/3 \quad C_g = C_A = 3$$

Infrared (IR) safety – observable is unchanged by adding a soft particle

Collinear (C) safety – observable is unchanged by a collinear splitting



(More in backup)

Infrared and Collinear (IRC) Safety

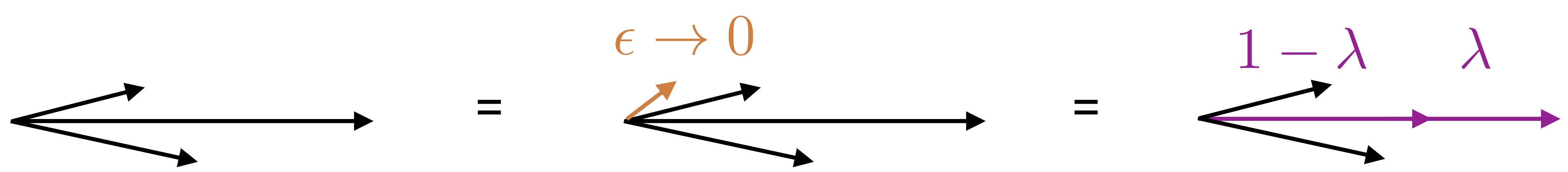
[**PTK**, Metodiev, Thaler, JHEP 2020]

QCD has soft and collinear divergences associated with gluon radiation

$$dP_{i\to ig} \simeq \frac{2\alpha_s}{\pi} C_a \frac{d\theta}{\theta} \frac{dz}{z} \qquad \begin{aligned} C_q &= C_F = 4/3 \\ C_g &= C_A = 3 \end{aligned}$$

Infrared (IR) safety – observable is unchanged by adding a soft particle

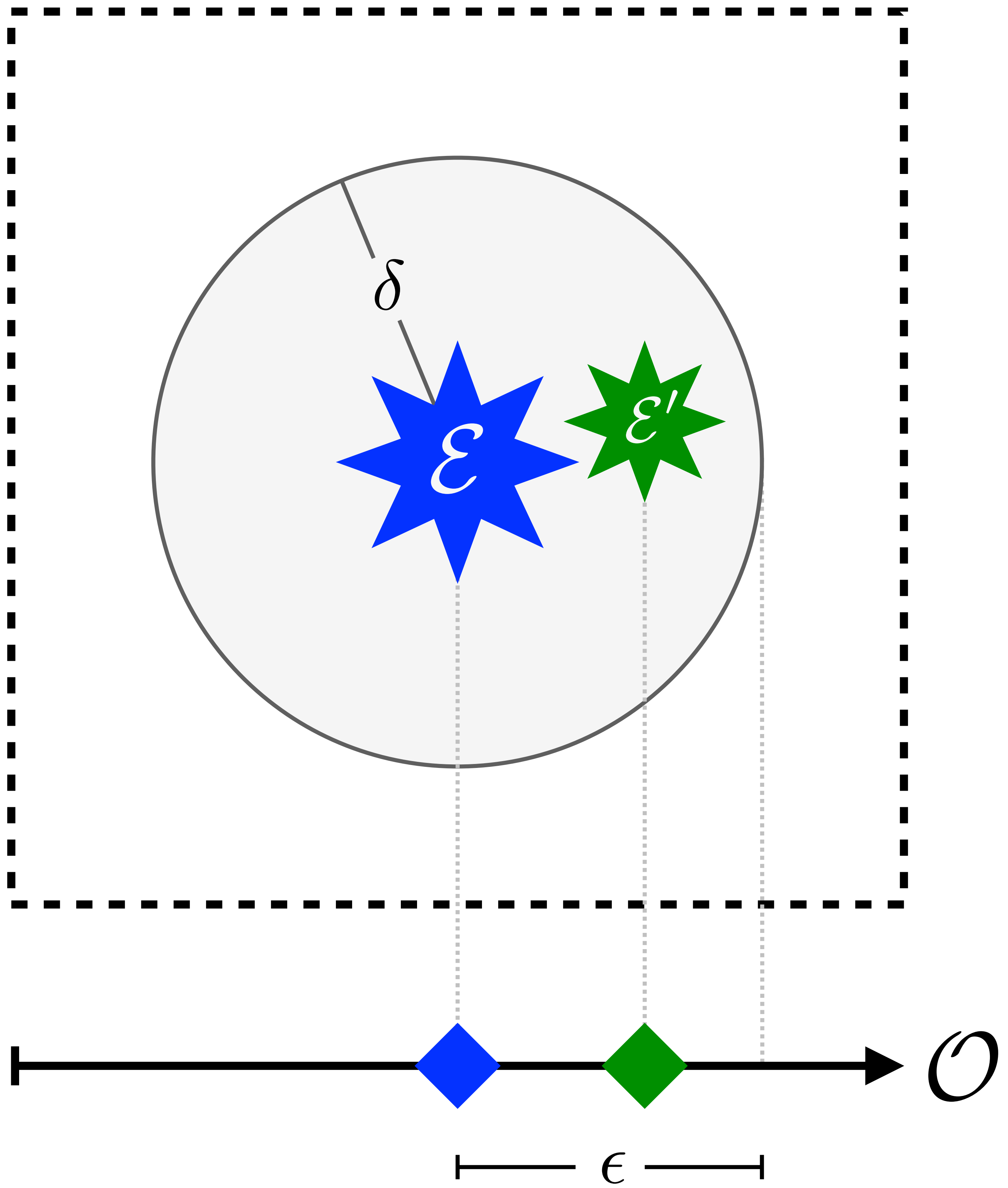
Collinear (C) safety – observable is unchanged by a collinear splitting



IRC Safety = EMD Continuity*

*on all but a negligible set‡ of events

‡a negligible set is one that contains no positive-radius EMD-ball



Classic $\epsilon - \delta$ definition of continuity in metric topology

An observable $\mathcal{O}$ is EMD continuous at an event $\mathcal{E}$ if, for any $\epsilon > 0$, there exists a $\delta > 0$ such that for all events $\mathcal{E}'$:

$$\mathrm{EMD}(\mathcal{E}, \mathcal{E}') < \delta \quad \implies \quad |\mathcal{O}(\mathcal{E}) - \mathcal{O}(\mathcal{E}')| < \epsilon.$$

(More in backup)

Energy Flow Methods

Energy Flow Polynomials (EFPs)

[**PTK**, Metodiev, Thaler, JHEP 2018]

Obtained via systematically expanding in energies and angles

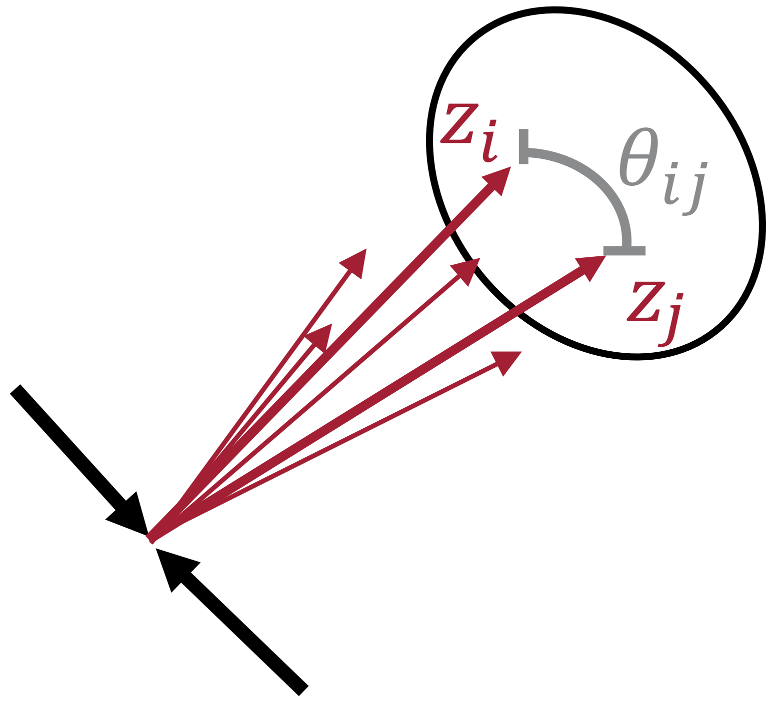
Definition of energy factor and pairwise angular distance

pp : $z_i = \frac{p_{Ti}}{\sum_j p_{Tj}}$ $\quad \theta_{ij}^2 = 2n_i^\mu n_{j\mu} = 2\frac{p_i^\mu}{p_{Ti}}\frac{p_{j\mu}}{p_{Tj}} \simeq (y_i - y_j)^2 + (\phi_i - \phi_j)^2$

e^+e^- : $z_i = \frac{E_i}{\sum_j E_j}$ $\quad \theta_{ij}^2 = 2n_i^\mu n_{j\mu} = 2\frac{p_i^\mu}{E_i}\frac{p_{j\mu}}{E_j}$

EFPs are *multiparticle correlators*

$$\sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} z_{i_1} \cdots z_{i_N} \prod_{(j,k)\in G} \theta_{i_j i_k}^{\beta}$$

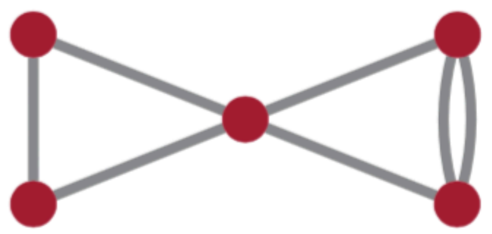


Graphs represent EFPs

vertex ↔ energy factor edge ↔ pairwise angle

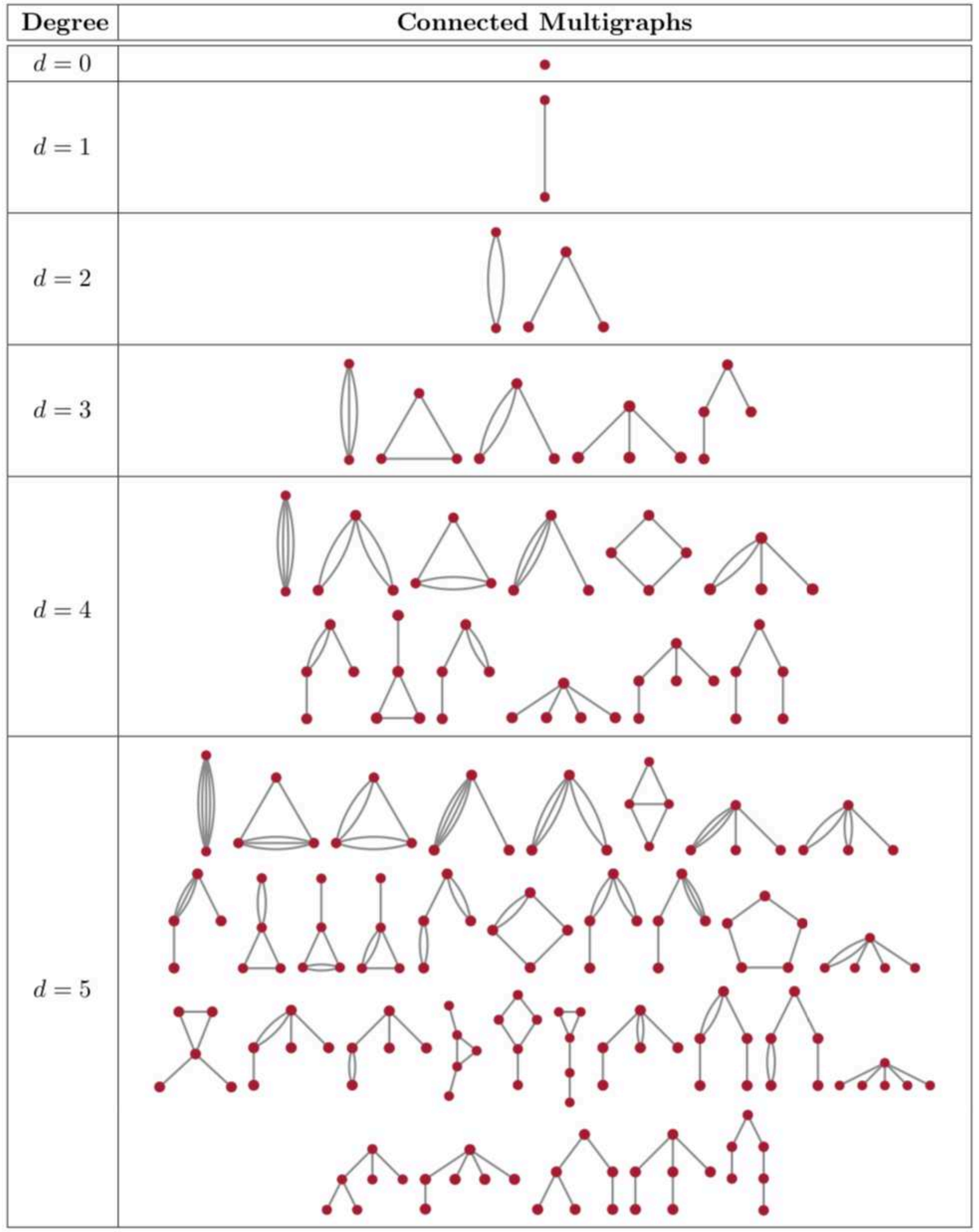
$j \longleftrightarrow \sum_{i_j=1}^{M} z_{i_j}$ $\quad k$ —— $l \longleftrightarrow \theta_{i_k i_l}$

e.g.



$$= \sum_{i_1=1}^{M}\sum_{i_2=1}^{M}\sum_{i_3=1}^{M}\sum_{i_4=1}^{M}\sum_{i_5=1}^{M} z_{i_1}z_{i_2}z_{i_3}z_{i_4}z_{i_5}\theta_{i_1i_2}\theta_{i_2i_3}\theta_{i_1i_3}\theta_{i_1i_4}\theta_{i_1i_5}\theta_{i_4i_5}^2$$

Organized by number of edges d



Any IRC-safe observable $\mathcal{S}$ is a linear combination of EFPs!

$$\mathcal{S} \simeq \sum_{G\in\mathcal{G}} s_G \mathrm{EFP}_G, \quad \mathcal{G} \text{ a set of multigraphs}$$

Energy Flow Moments (EFMs)

[**PTK**, Metodiev, Thaler, PRD 2020]

$\theta_{ij} = \sqrt{2 n_i^\mu n_{j\mu}}$ $\beta = 2$ removes square root

Factors of n_i^μ can be organized in optimal way

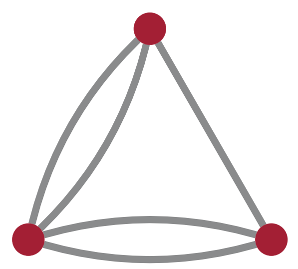
EFM$_v$ is a little group tensor with v indices

$$\mathcal{I}^{\mu_1 \cdots \mu_v} = \sum_{i=1}^{M} z_i n_i^{\mu_1} \cdots n_i^{\mu_v}$$

v	0	1	2	3	4	5	6
$n_{\text{components}}^{(d=4)}$	1	4	10	20	35	56	84

$$\mathcal{I}^{j_1 j_2 \cdots j_v} = 2^{v/2} \Theta^{j_1 j_2 \cdots j_v}$$

spatial e^+e^- EFMs — linearized sphericity tensors [Donoghue, Low, Pi, PRD 1979]



$$= \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} \sum_{i_3=1}^{M} z_{i_1} z_{i_2} z_{i_3} \theta_{i_1 i_2}^2 \theta_{i_1 i_3}^2 \theta_{i_2 i_3}$$

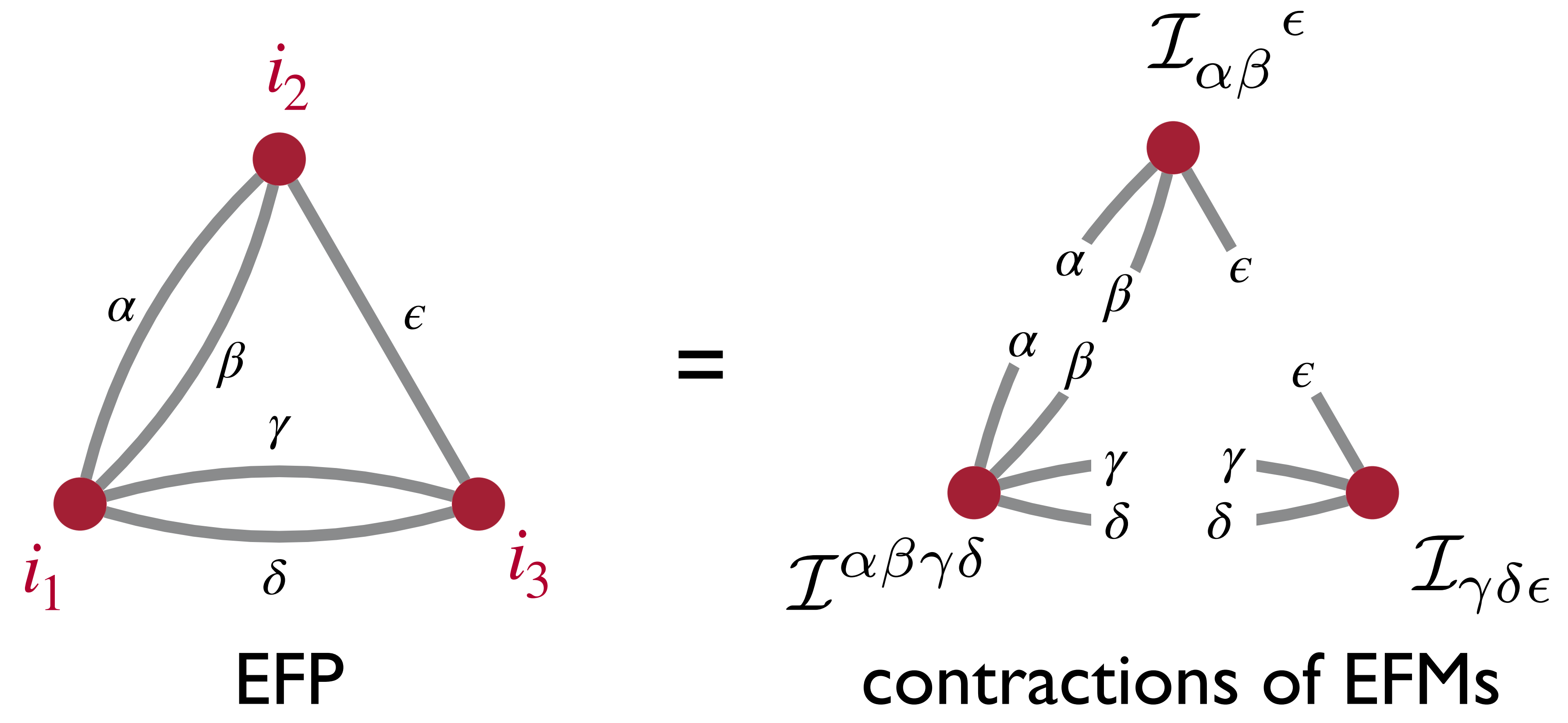
$$= 2^5 \underbrace{\left(\sum_{i_1=1}^{M} z_{i_1} n_{i_1}^{\alpha} n_{i_1}^{\beta} n_{i_1}^{\gamma} n_{i_1}^{\delta} \right)}_{\mathcal{I}^{\alpha\beta\gamma\delta}} \underbrace{\left(\sum_{i_2=1}^{M} z_{i_2} n_{i_2\alpha} n_{i_2\beta} n_{i_2}^{\epsilon} \right)}_{\mathcal{I}_{\alpha\beta}{}^{\epsilon}} \underbrace{\left(\sum_{i_3=1}^{M} z_{i_3} n_{i_3\gamma} n_{i_3\delta} n_{i_3\epsilon} \right)}_{\mathcal{I}_{\gamma\delta\epsilon}}$$

Naively $\mathcal{O}(M^3)$ EFP shown to be $\mathcal{O}(M)$

All $\beta = 2$ EFPs are $\mathcal{O}(M)$

- $\text{ECF}_N^{(\beta=2)}$ are all $\mathcal{O}(M)$
- $D_2^{(\beta=2)}$, $C_2^{(\beta=2)}$ are $\mathcal{O}(M)$

EFMs result from cutting edges of EFP



(See backup for more on understanding linear redundancies and counting superstring amplitudes with EFMs)

Understanding Linear Redundancies via EFMs

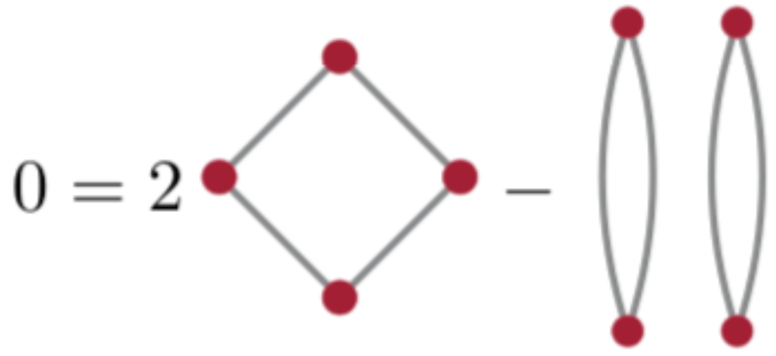
[PTK, Metodiev, Thaler, PRD 2020]

Linear redundancies among EFPs are troublesome

Studying coefficients of linear fit difficult $\mathcal{O} = \sum_G s_G \mathrm{EFP}_G$

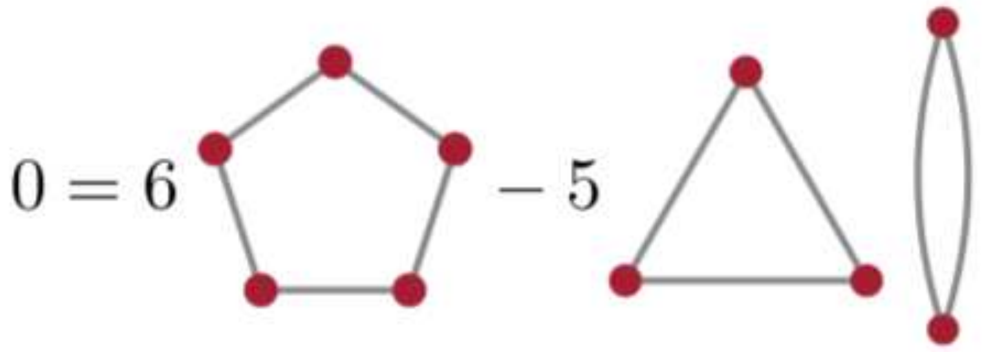
Examples of redundancies

in 3 or fewer spacetime dimensions

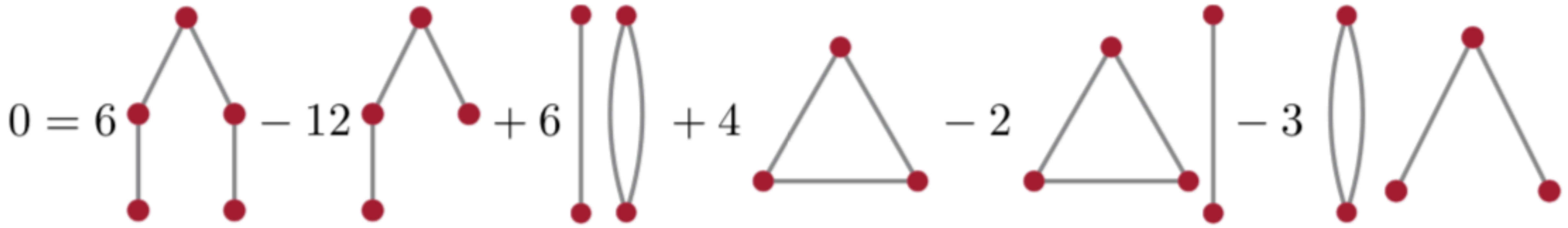


$$0 = \mathcal{I}^{\beta}_{[\alpha}\mathcal{I}^{\gamma}_{\beta}\mathcal{I}^{\delta}_{\gamma}\mathcal{I}^{\alpha}_{\delta]}$$

in 4 or fewer spacetime dimensions



$$0 = \mathcal{I}^{\beta}_{[\alpha}\mathcal{I}^{\gamma}_{\beta}\mathcal{I}^{\delta}_{\gamma}\mathcal{I}^{\epsilon}_{\delta}\mathcal{I}^{\alpha}_{\epsilon]}$$



How to obtain a tensor identity

Consider tensor over *n* dimensional vector space

Antisymmetrize *m* > *n* indices

Result is zero because any assignment of *n* possible values to *m* slots has a repetition

$$T^{a_1 \cdots a_k}_{b_1 \cdots b_\ell [c_1 \cdots c_m]} = 0$$

Bonus: all tensor identities up to ones governed by existing symmetries take above form

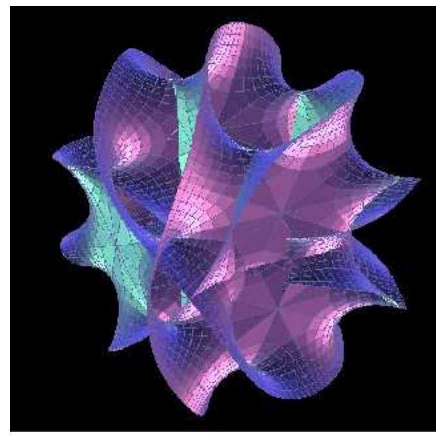
[Sneddon, Journal of Mathematical Physics]

In e^+e^- there are additional relations due to

$$n_i^\mu = (1, \hat{n})^\mu \implies \mathcal{I}^{0\mu_1\cdots\mu_v} = \sqrt{2}\mathcal{I}^{\mu_1\cdots\mu_v}$$

See backup for more on these "Euclidean" relations

Counting Superstring Amplitudes



Constructing a basis of amplitudes – how large is it?

[Boels, 1304.7918; OEIS A226919]

non-isomorphic multigraph

$\Updownarrow$

Q: What is the number of symmetric polynomials of degree d in kinematic variables up to momentum conservation?

$\Updownarrow$ $\Updownarrow$

$$\theta_{ij}^2 = 2n_i \cdot n_j$$

pairwise angular distance

$$0 = \sum_{i=1}^{M} p_i^\mu = \mathcal{I}^\mu =$$ 

A: Same as the number of non-isomorphic multigraphs with no leaves (vertices of valency one)

New OEIS Entries!
A307317, A307316

	Leafless Multigraphs	
	Connected	All
Edges d	A307317	A307316
1	0	0
2	1	1
3	2	2
4	4	5
5	9	11
6	26	34
7	68	87
8	217	279
9	**718**	**897**
10	**2 553**	**3 129**
11	**9 574**	**11 458**
12	**38 005**	**44 576**
13	**157 306**	**181 071**
14	**679 682**	**770 237**
15	**3 047 699**	**3 407 332**
16	**14 150 278**	**15 641 159**

Bolded values previously unknown

Machine Learning for Point Clouds – Deep Sets

A general permutation-symmetric function is additive in a latent space

Deep Sets

[1703.06114]

Manzil Zaheer[1,2], **Satwik Kottur**[1], **Siamak Ravanbhakhsh**[1],
Barnabás Póczos[1], **Ruslan Salakhutdinov**[1], **Alexander J Smola**[1,2]
[1] Carnegie Mellon University [2] Amazon Web Services

Deep Sets Theorem [63]. *Let $\mathfrak{X} \subset \mathbb{R}^d$ be compact, $X \subset 2^{\mathfrak{X}}$ be the space of sets with bounded cardinality of elements in $\mathfrak{X}$, and $Y \subset \mathbb{R}$ be a bounded interval. Consider a continuous function $f : X \to Y$ that is invariant under permutations of its inputs, i.e. $f(x_1, \dots, x_M) = f(x_{\pi(1)}, \dots, x_{\pi(M)})$ for all $x_i \in \mathfrak{X}$ and $\pi \in S_M$. Then there exists a sufficiently large integer ℓ and continuous functions $\Phi : \mathfrak{X} \to \mathbb{R}^\ell$, $F : \mathbb{R}^\ell \to Y$ such that the following holds to an arbitrarily good approximation:*[1]

$$f(\{x_1, \dots, x_M\}) = F\left(\sum_{i=1}^{M} \Phi(x_i)\right)$$

Machine Learning for Point Clouds – Deep Sets

A general permutation-symmetric function is additive in a latent space

Deep Sets

[1703.06114]

Manzil Zaheer[1,2], **Satwik Kottur**[1], **Siamak Ravanbhakhsh**[1], **Barnabás Póczos**[1], **Ruslan Salakhutdinov**[1], **Alexander J Smola**[1,2]
[1] Carnegie Mellon University [2] Amazon Web Services

Feature space

Variable length

Deep Sets Theorem [63]. *Let $\mathfrak{X} \subset \mathbb{R}^d$ be compact, $X \subset 2^{\mathfrak{X}}$ be the space of sets with bounded cardinality of elements in $\mathfrak{X}$, and $Y \subset \mathbb{R}$ be a bounded interval. Consider a continuous function $f : X \to Y$ that is invariant under permutations of its inputs, i.e. $f(x_1, \ldots, x_M) = f(x_{\pi(1)}, \ldots, x_{\pi(M)})$ for all $x_i \in \mathfrak{X}$ and $\pi \in S_M$. Then there exists a sufficiently large integer ℓ and continuous functions $\Phi : \mathfrak{X} \to \mathbb{R}^\ell$, $F : \mathbb{R}^\ell \to Y$ such that the following holds to an arbitrarily good approximation:*[1]

Permutation invariance

Latent space

$$f(\{x_1, \ldots, x_M\}) = F\left(\sum_{i=1}^{M} \Phi(x_i)\right)$$

General parametrization for a function of sets

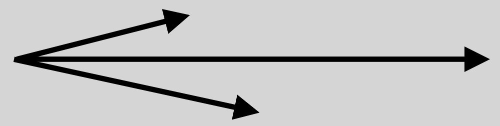
Infrared and Collinear (IRC) Safety

QCD has soft and collinear divergences associated with gluon radiation

dP_{i-}

Infrared (IR) safety – o

Collinear (C) safety – c



=

IRC safety is a statement of *linearity* in energy and *continuity* in geometry

Theorem: Any IRC-safe observable can be written in the following form:

$$f(\{p_1^\mu,\ldots,p_M^\mu\}) = F\left(\sum_{i=1}^{M} z_i \vec{\Phi}(\hat{p}_i)\right), \quad \hat{p}_i = (y_i,\phi_i).$$

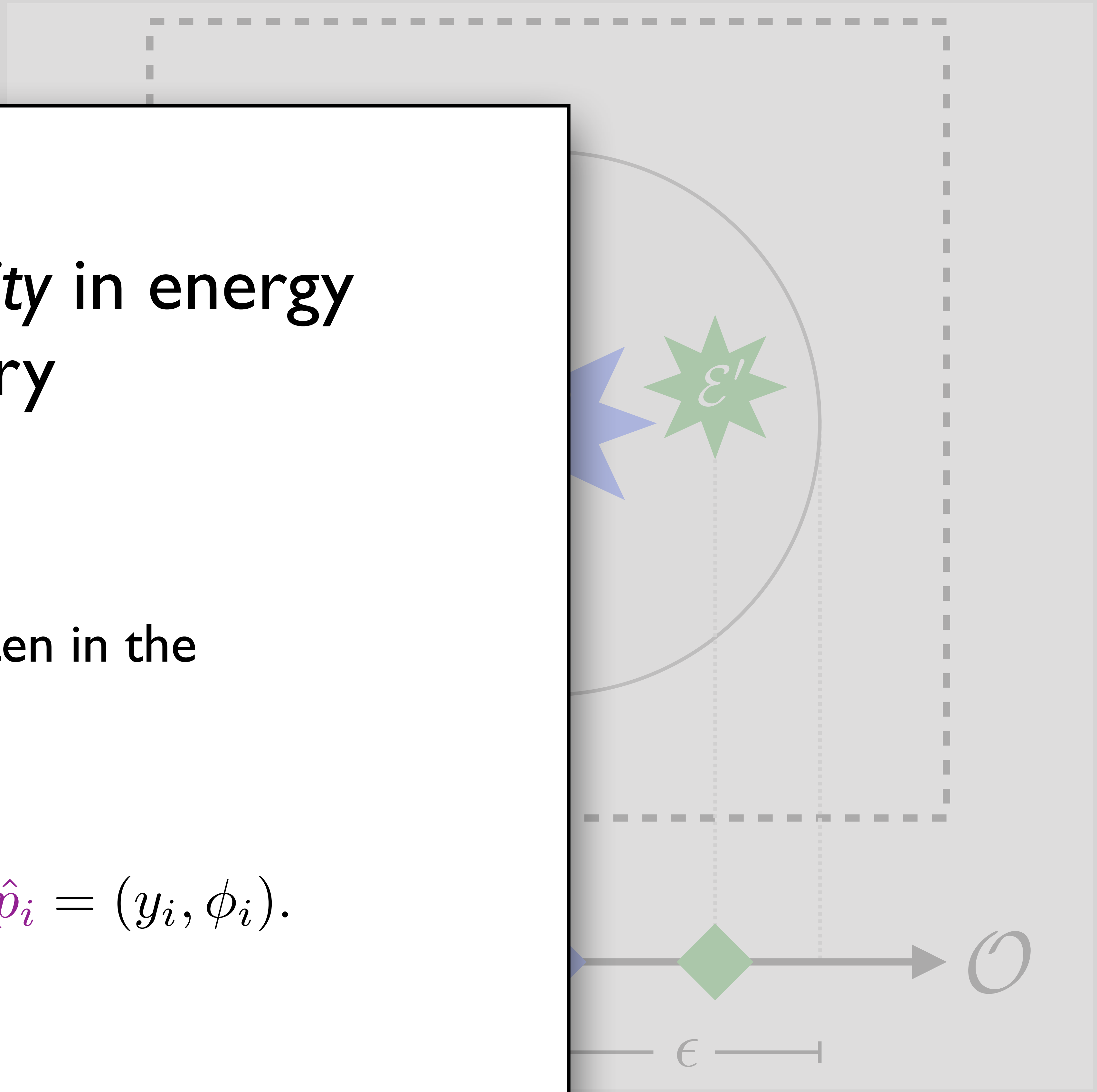
Proof: In 1810.05165. □

Infrared and Collinear (IRC) Safety

[**PTK**, Metodiev, Thaler, JHEP 2020]

QCD has soft and collinear divergences associated with gluon radiation

dP_{i}



IRC safety is a statement of *linearity* in energy and *continuity* in geometry

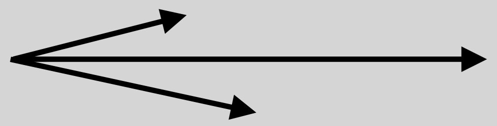
Theorem: Any IRC-safe observable can be written in the following form:

$$f(\{p_1^\mu, \ldots, p_M^\mu\}) = F\left(\sum_{i=1}^{M} z_i \vec{\Phi}(\hat{p}_i)\right), \quad \hat{p}_i = (y_i, \phi_i).$$

Proof: In 1810.05165. □

Infrared (IR) safety – ol

Collinear (C) safety – c



=

IRC Safe

*on all but a negligible set‡ of events

‡a negligible set is one that contains no positive-radius EMD-ball

ntinuity in metric topology

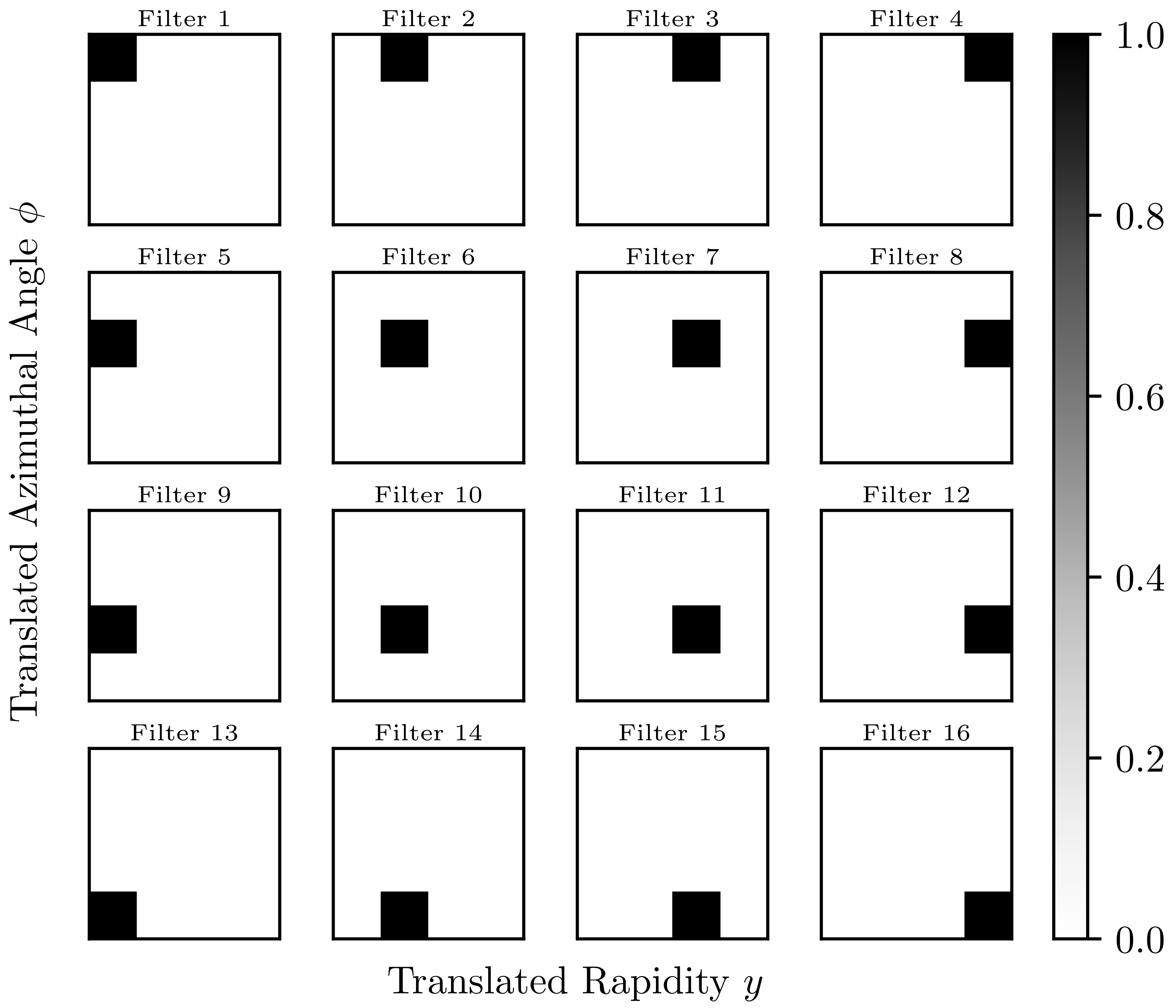
tinuous at an event $\mathcal{E}$ if, for any $\epsilon > 0$, there exists a $\delta > 0$ such that for all events $\mathcal{E}'$:

$$\text{EMD}(\mathcal{E}, \mathcal{E}') < \delta \quad \Longrightarrow \quad |\mathcal{O}(\mathcal{E}) - \mathcal{O}(\mathcal{E}')| < \epsilon.$$

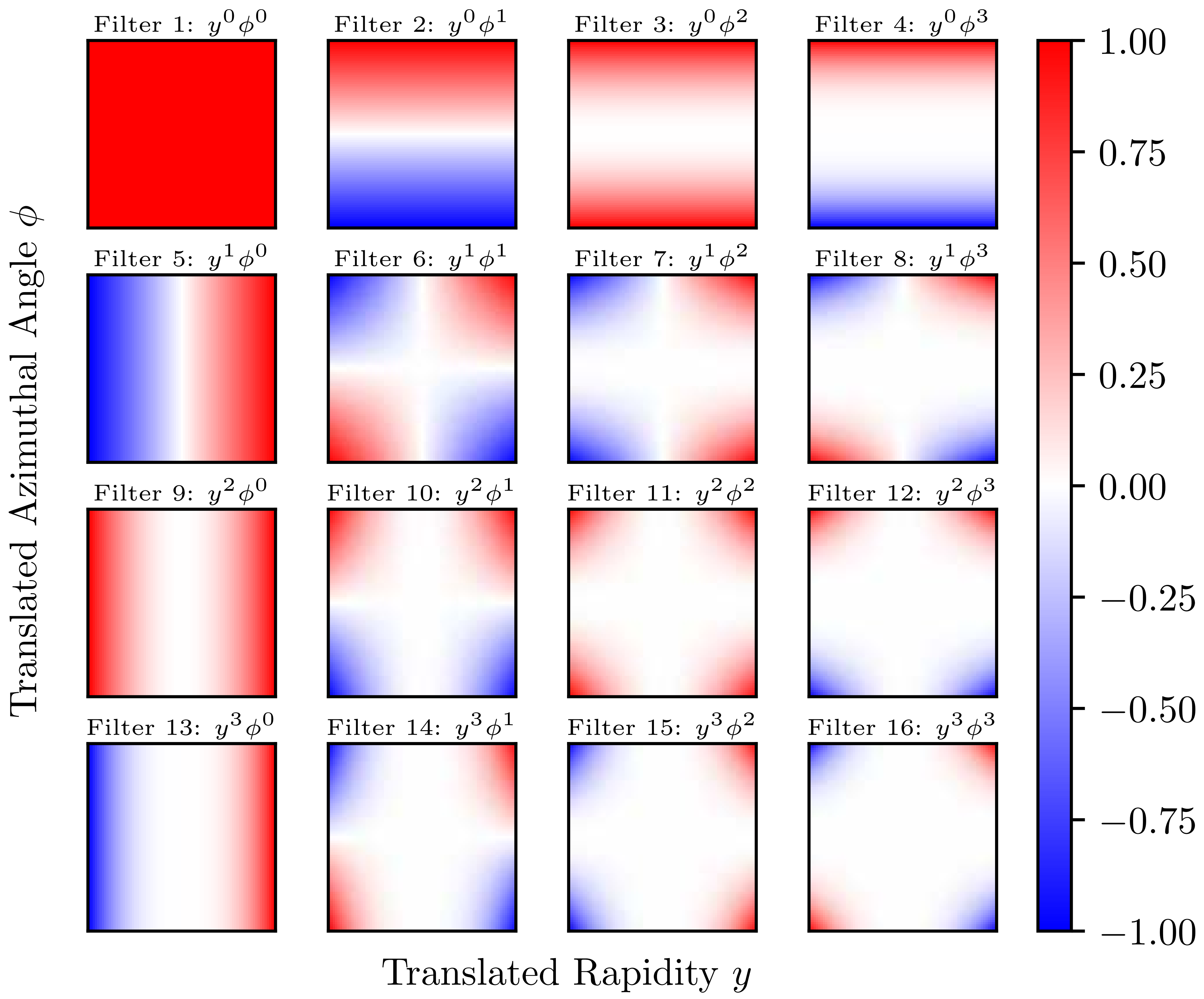
Energy Flow Network Visualization

Visualize EFN observables as 2D filters in the translated rapidity-azimuth plane

Jet images as EFN filters



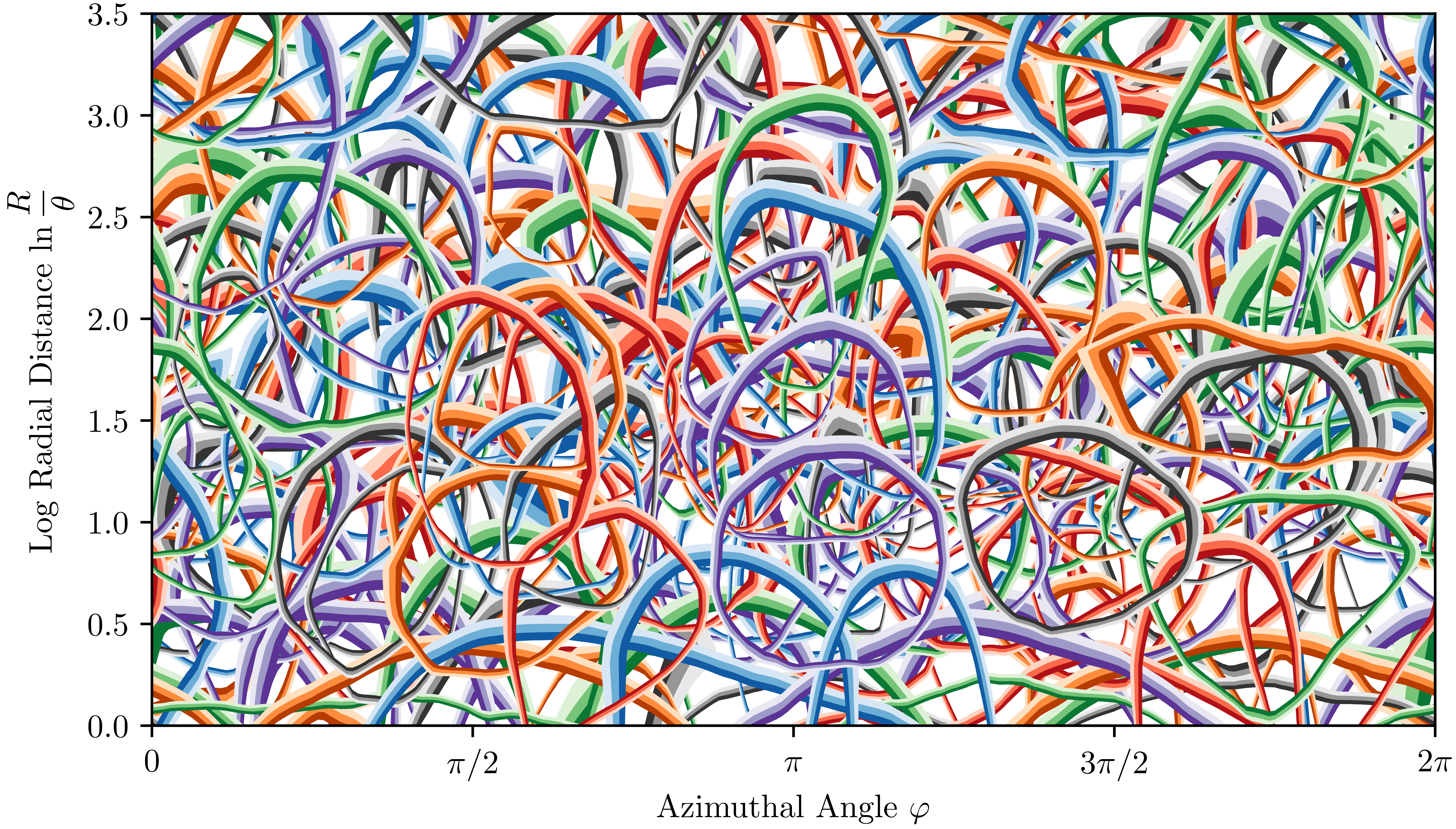
Moments as EFN filters



[Donoghue, Low, Pi, PRD 1979; Gur-Ari, Papucci, Perez, 1101.2905;
PTK, Metodiev, Thaler, PRD 2020]

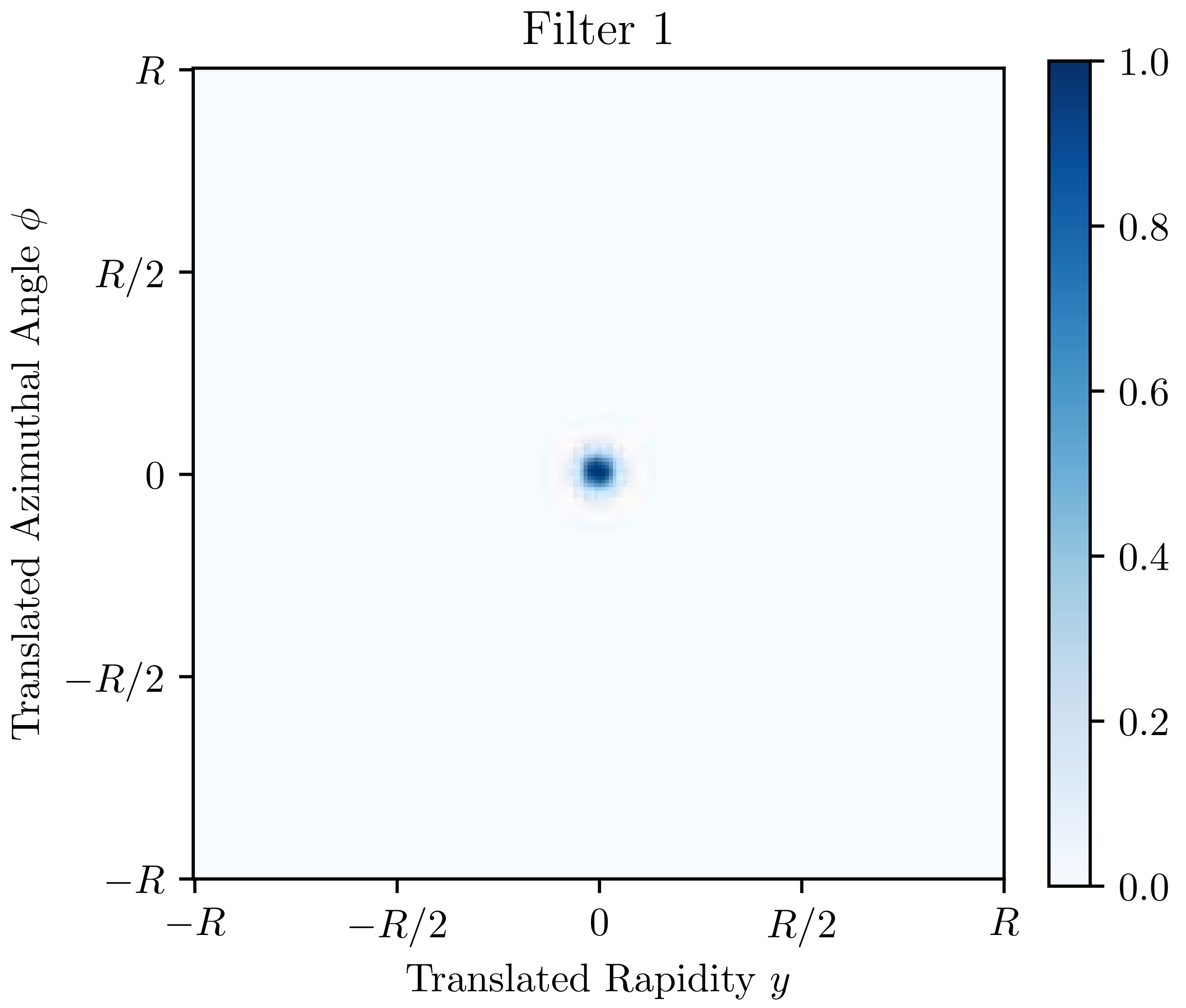
Quark vs. Gluon: Visualizing EFN Filters in the Emission Plane

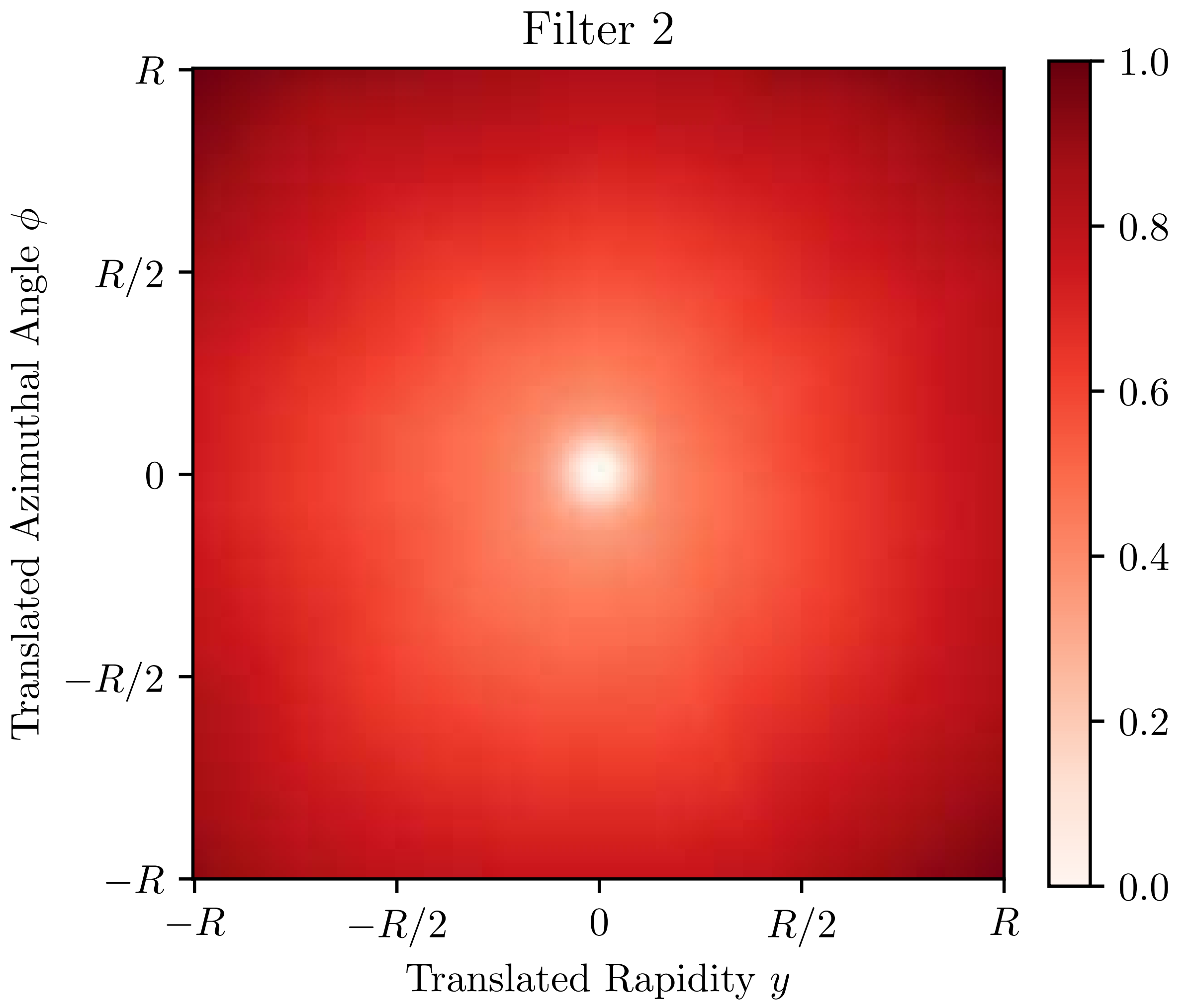
Transform to polar coordinates and take logarithm of the radius



Quark vs. Gluon: Extracting New Analytic Observables

EFN (ℓ = 2) has approximately radially symmetric filters

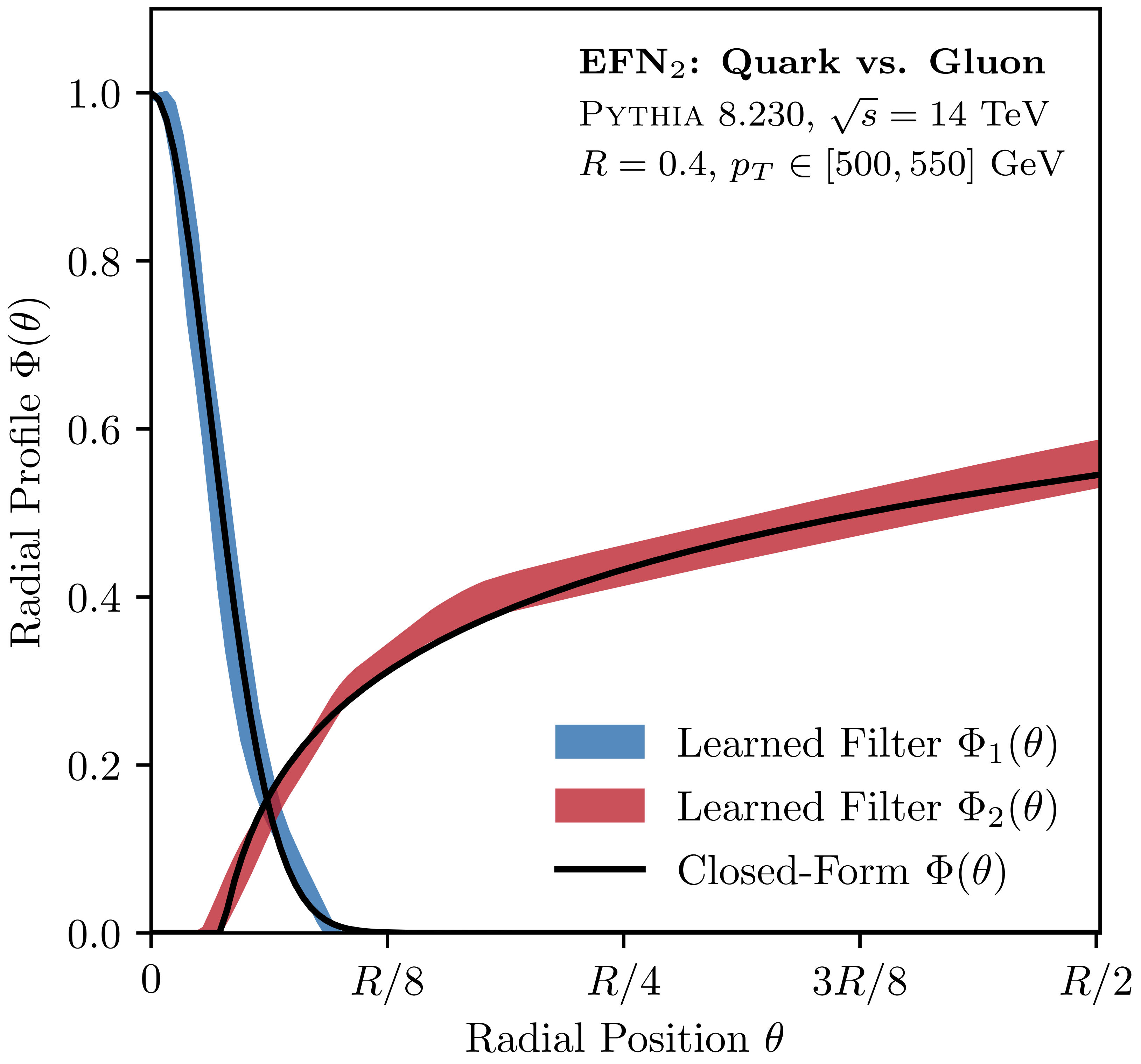




$$\mathcal{O}_1 = \sum_{i=1}^{M} z_i \, \Phi_1(\theta_i)$$

$$\mathcal{O}_2 = \sum_{i=1}^{M} z_i \, \Phi_2(\theta_i)$$

Average of radial slices



Separate soft and collinear phase space regions,
e.g. collinear drop

[Chien, Stewart, JHEP 2020]

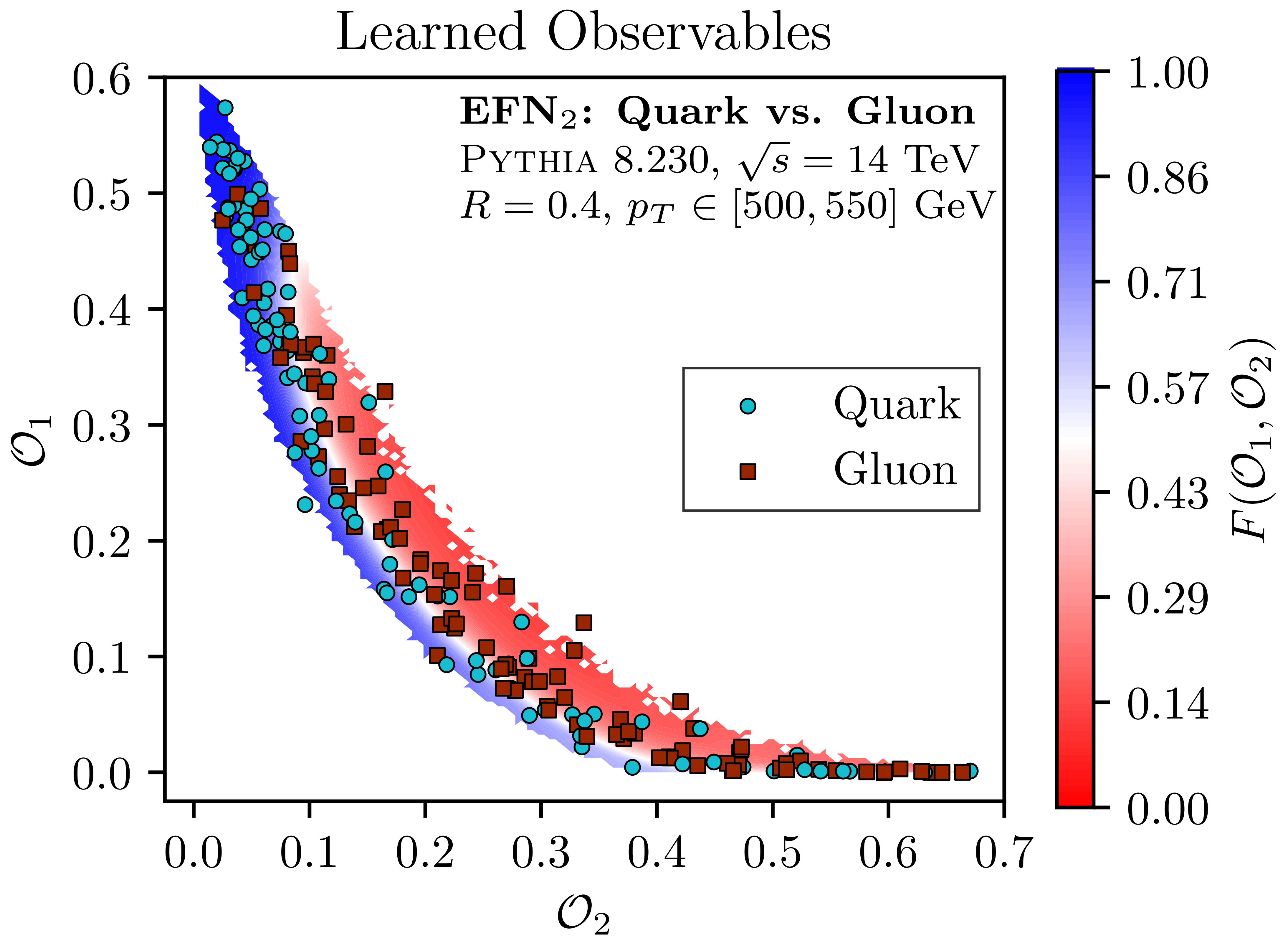
Fit analytic forms:

$$A_{r_0} = \sum_{i=1}^{M} z_i \, e^{-\theta_i^2/r_0^2},$$

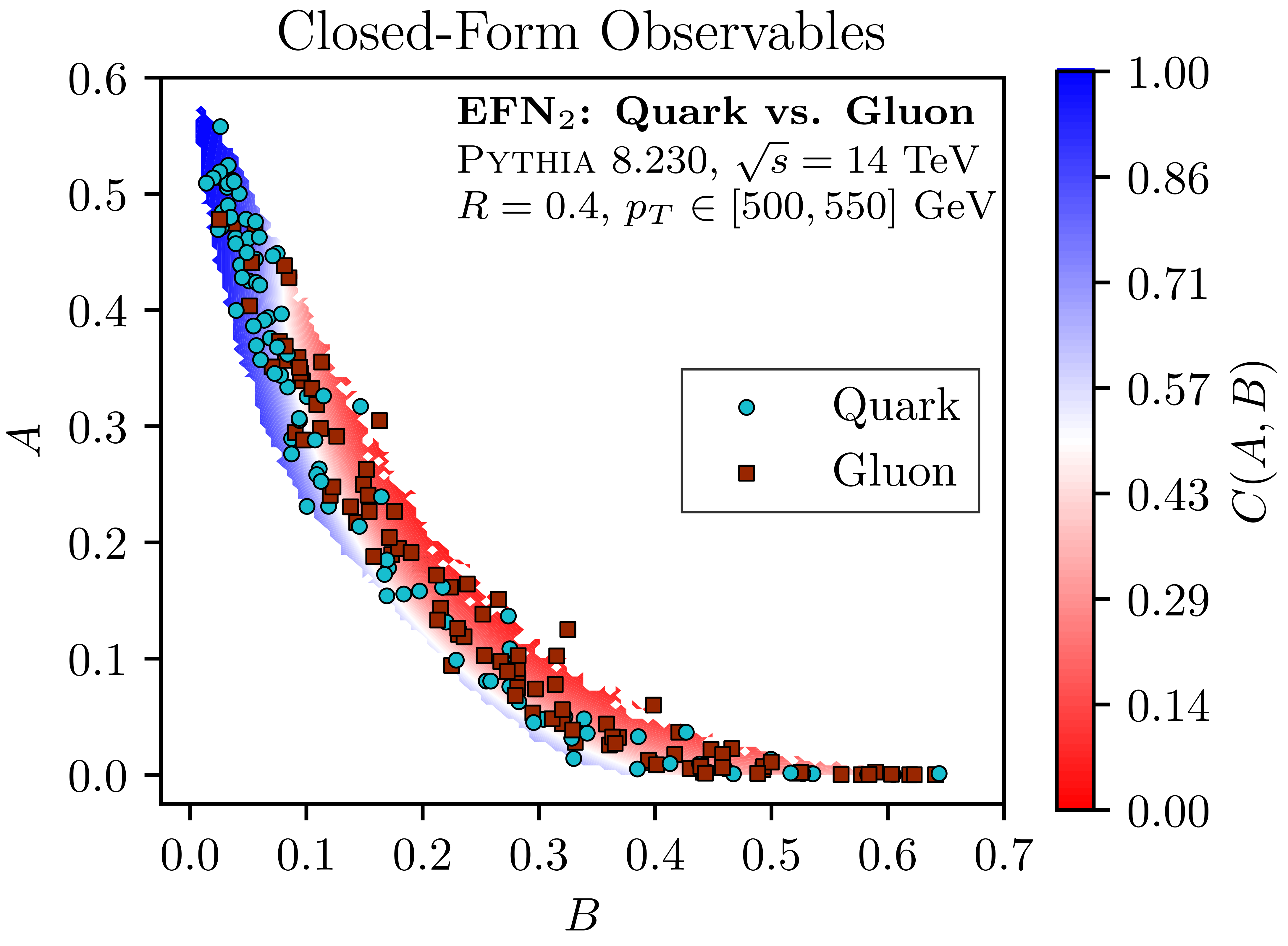
$$B_{r_1,\beta} = \sum_{i=1}^{M} z_i \, \ln(1 + \beta(\theta_i - r_1))\Theta(\theta_i - r_1)$$

Quark vs. Gluon: Extracting New Analytic Observables

Visualize F in the two dimensional $(\mathcal{O}_1, \mathcal{O}_2)$ phase space



Learned with EFN (ℓ=2)



Analytic Fit

Squared distance from a point:

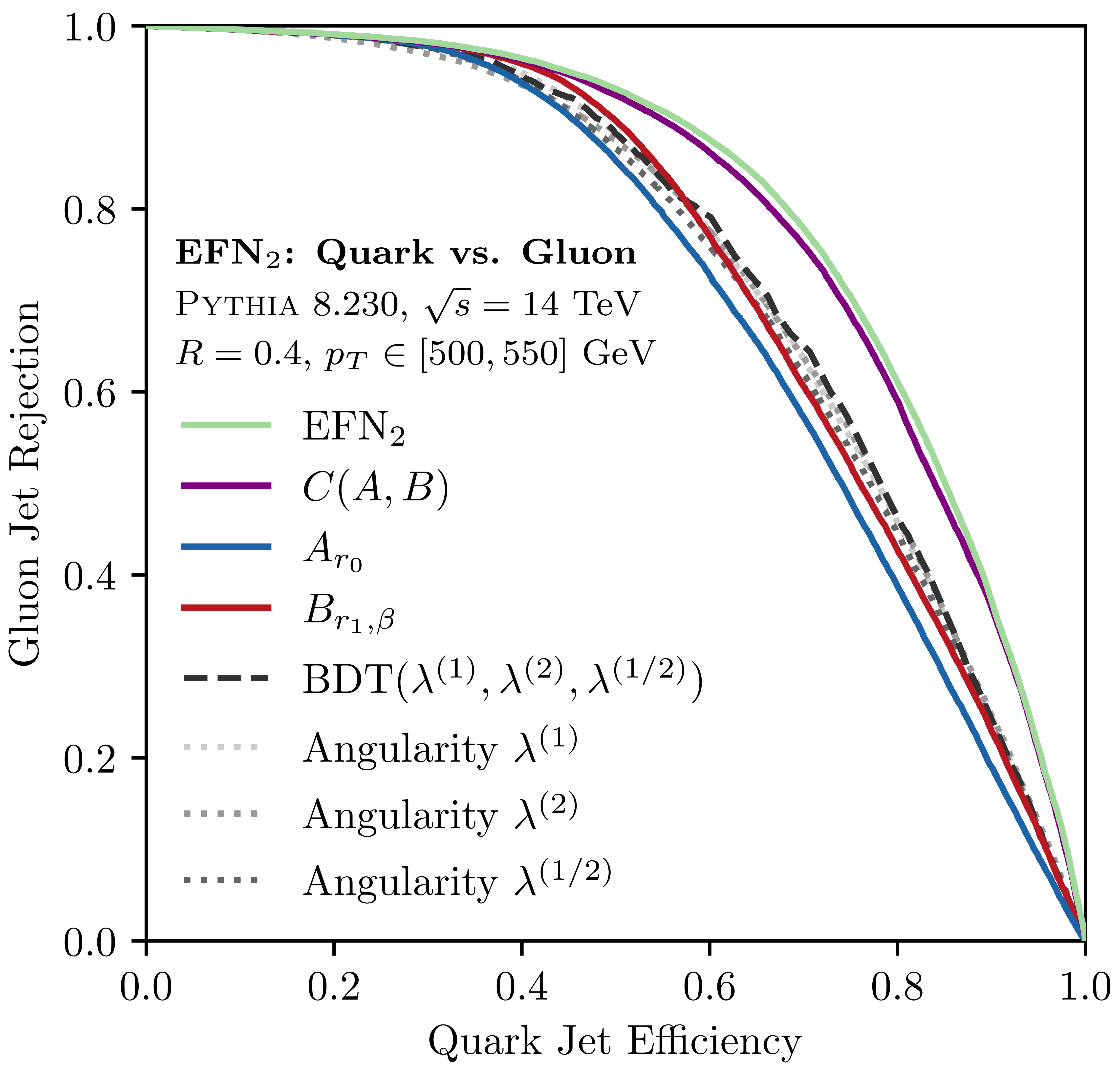
$$C(A, B) = (A - a_0)^2 + (B - b_0)^2$$

Quark vs. Gluon: Benchmarking New Analytic Observables

A, *B* observables individually comparable to angularities

$C(A,B)$ vastly exceeds multivariate combination (BDT) of angularities

Extracted $C(A,B)$ is comparable to EFN ($\ell = 2$)



OmniFold

Iterated Bayesian Unfolding (IBU)

[Richardson, JOSA 1972; Lucy, AJ 1974; D'Agostini, NIMPA 1995]

Maximum likelihood, histogram-based unfolding method for a small number of observables

Choose observable(s) and binning at detector-level and particle-level

measured distribution: $m_i = \Pr(\text{measure } i)$

true distribution: $t_j^{(0)} = \Pr(\text{truth is } j)$

Calculate *response matrix* R_{ij} from generated/simulated pairs of events

$$R_{ij} = \Pr(\text{measure } i \mid \text{truth is } j)$$

Calculate new particle-level distribution using Bayes' theorem

$$t_j^{(n)} = \sum_i \Pr(\text{truth}_{n-1} \text{ is } j \mid \text{measure } i) \times \Pr(\text{measure } i)$$
$$= \sum_i \frac{R_{ij} t_j^{(n-1)}}{\sum_k R_{ik} t_k^{(n-1)}} \times m_i$$

Iterate procedure to remove dependence on prior

Consider a situation with two particle-level bins and two detector-level bins

$$t_j^{(0)} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_j \qquad m_i = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i \qquad R_{ij} = \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{pmatrix}_{ij}$$

Uniform prior

Bins are measured equally

Bin 1 reconstructed perfectly
Bin 2 reconstructed equally

Iterated Bayesian Unfolding (IBU)

[Richardson, JOSA 1972; Lucy, AJ 1974; D'Agostini, NIMPA 1995]

Maximum likelihood, histogram-based unfolding method for a small number of observables

Choose observable(s) and binning at detector-level and particle-level

measured distribution: $m_i = \Pr(\text{measure } i)$

true distribution: $t_j^{(0)} = \Pr(\text{truth is } j)$

Calculate *response matrix* R_{ij} from generated/simulated pairs of events

$$R_{ij} = \Pr(\text{measure } i \mid \text{truth is } j)$$

Calculate new particle-level distribution using Bayes' theorem

$$t_j^{(n)} = \sum_i \Pr(\text{truth}_{n-1} \text{ is } j \mid \text{measure } i) \times \Pr(\text{measure } i) = \sum_i \frac{R_{ij} t_j^{(n-1)}}{\sum_k R_{ik} t_k^{(n-1)}} \times m_i$$

Iterate procedure to remove dependence on prior

Consider a situation with two particle-level bins and two detector-level bins

$$t_j^{(0)} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_j \qquad m_i = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i \qquad R_{ij} = \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{pmatrix}_{ij}$$

Uniform prior

Bins are measured equally

Bin 1 reconstructed perfectly
Bin 2 reconstructed equally

$$t_j^{(1)} = \sum_i \frac{\begin{pmatrix} \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{4} \end{pmatrix}_{ij}}{\begin{pmatrix} \frac{3}{4} & \frac{1}{4} \end{pmatrix}_i} \times \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i$$

After one iteration

Iterated Bayesian Unfolding (IBU)

[Richardson, JOSA 1972; Lucy, AJ 1974; D'Agostini, NIMPA 1995]

Maximum likelihood, histogram-based unfolding method for a small number of observables

Choose observable(s) and binning at detector-level and particle-level

measured distribution: $m_i = \Pr(\text{measure } i)$

true distribution: $t_j^{(0)} = \Pr(\text{truth is } j)$

Calculate *response matrix* R_{ij} from generated/simulated pairs of events

$$R_{ij} = \Pr(\text{measure } i \mid \text{truth is } j)$$

Calculate new particle-level distribution using Bayes' theorem

$$t_j^{(n)} = \sum_i \Pr(\text{truth}_{n-1} \text{ is } j \mid \text{measure } i) \times \Pr(\text{measure } i) = \sum_i \frac{R_{ij} t_j^{(n-1)}}{\sum_k R_{ik} t_k^{(n-1)}} \times m_i$$

Iterate procedure to remove dependence on prior

Consider a situation with two particle-level bins and two detector-level bins

$$t_j^{(0)} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_j \qquad m_i = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i \qquad R_{ij} = \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{pmatrix}_{ij}$$

Uniform prior

Bins are measured equally

Bin 1 reconstructed perfectly
Bin 2 reconstructed equally

$$t_j^{(1)} = \sum_i \frac{\begin{pmatrix} \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{4} \end{pmatrix}_{ij}}{\begin{pmatrix} \frac{3}{4} & \frac{1}{4} \end{pmatrix}_i} \times \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i = \begin{pmatrix} \underbrace{\frac{1}{2} \times \frac{4}{3} \times \frac{1}{2}}_{1/3} + 0 \\ \underbrace{\frac{1}{2} \times \frac{4}{3} \times \frac{1}{4}}_{1/6} + \underbrace{\frac{1}{2} \times 4 \times \frac{1}{4}}_{1/2} \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \end{pmatrix}_j$$

After one iteration

Iterated Bayesian Unfolding (IBU)

[Richardson, JOSA 1972; Lucy, AJ 1974; D'Agostini, NIMPA 1995]

Maximum likelihood, histogram-based unfolding method for a small number of observables

Choose observable(s) and binning at detector-level and particle-level

measured distribution: $m_i = \Pr(\text{measure } i)$

true distribution: $t_j^{(0)} = \Pr(\text{truth is } j)$

Calculate *response matrix* R_{ij} from generated/simulated pairs of events

$$R_{ij} = \Pr(\text{measure } i \mid \text{truth is } j)$$

Calculate new particle-level distribution using Bayes' theorem

$$t_j^{(n)} = \sum_i \Pr(\text{truth}_{n-1} \text{ is } j \mid \text{measure } i) \times \Pr(\text{measure } i) = \sum_i \frac{R_{ij} t_j^{(n-1)}}{\sum_k R_{ik} t_k^{(n-1)}} \times m_i$$

Iterate procedure to remove dependence on prior

Consider a situation with two particle-level bins and two detector-level bins

$$t_j^{(0)} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_j \qquad m_i = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i \qquad R_{ij} = \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{pmatrix}_{ij}$$

Uniform prior

Bins are measured equally

Bin 1 reconstructed perfectly
Bin 2 reconstructed equally

$$t_j^{(1)} = \sum_i \frac{\begin{pmatrix} \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{4} \end{pmatrix}_{ij}}{\begin{pmatrix} \frac{3}{4} & \frac{1}{4} \end{pmatrix}_i} \times \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i = \begin{pmatrix} \underbrace{\frac{1}{2} \times \frac{4}{3} \times \frac{1}{2}}_{1/3} + 0 \\ \underbrace{\frac{1}{2} \times \frac{4}{3} \times \frac{1}{4}}_{1/6} + \underbrace{\frac{1}{2} \times 4 \times \frac{1}{4}}_{1/2} \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \end{pmatrix}_j$$

After one iteration

⋮

$$t_j^{(n)} = \sum_i \frac{\begin{pmatrix} \frac{1}{n+1} & \frac{n}{2(n+1)} \\ 0 & \frac{n}{2(n+1)} \end{pmatrix}_{ij}}{\begin{pmatrix} \frac{n+2}{2(n+1)} & \frac{n}{2(n+1)} \end{pmatrix}_i} \times \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i = \begin{pmatrix} \frac{1}{n+2} \\ \frac{n+1}{n+2} \end{pmatrix}_j \to \begin{pmatrix} 0 \\ 1 \end{pmatrix}_j$$

Correct truth distribution obtained as $n \to \infty$

At the n[th] iteration

IBU as Reweighting

[Richardson, JOSA 1972; Lucy, AJ 1974; D'Agostini, NIMPA 1995]

Consider a situation with two particle-level bins and two detector-level bins

$$t_j^{(0)} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_j \qquad m_i = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i \qquad R_{ij} = \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{pmatrix}_{ij}$$

Uniform prior

Bins are measured equally

Bin 1 reconstructed perfectly
Bin 2 reconstructed equally

$$t_j^{(1)} = \sum_i \frac{\begin{pmatrix} \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{4} \end{pmatrix}_{ij}}{\begin{pmatrix} \frac{3}{4} & \frac{1}{4} \end{pmatrix}_i} \times \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i = \begin{pmatrix} \underbrace{\frac{1}{2} \times \frac{4}{3} \times \frac{1}{2}}_{1/3} + 0 \\ \underbrace{\frac{1}{2} \times \frac{4}{3} \times \frac{1}{4}}_{1/6} + \underbrace{\frac{1}{2} \times 4 \times \frac{1}{4}}_{1/2} \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \end{pmatrix}_j$$

After one iteration

$\vdots$

$$t_j^{(n)} = \sum_i \frac{\begin{pmatrix} \frac{1}{n+1} & \frac{n}{2(n+1)} \\ 0 & \frac{n}{2(n+1)} \end{pmatrix}_{ij}}{\begin{pmatrix} \frac{n+2}{2(n+1)} & \frac{n}{2(n+1)} \end{pmatrix}_i} \times \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i = \begin{pmatrix} \frac{1}{n+2} \\ \frac{n+1}{n+2} \end{pmatrix}_j \to \begin{pmatrix} 0 \\ 1 \end{pmatrix}_j$$

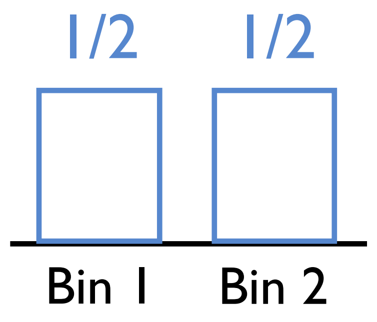
Correct truth distribution obtained as $n \to \infty$

At the n^{th} iteration

IBU as Reweighting

[Richardson, JOSA 1972; Lucy, AJ 1974; D'Agostini, NIMPA 1995]

prior



Consider a situation with two particle-level bins and two detector-level bins

$$t_j^{(0)} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_j \qquad m_i = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i \qquad R_{ij} = \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{pmatrix}_{ij}$$

Uniform prior

Bins are measured equally

Bin 1 reconstructed perfectly
Bin 2 reconstructed equally

$$t_j^{(1)} = \sum_i \frac{\begin{pmatrix} \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{4} \end{pmatrix}_{ij}}{\begin{pmatrix} \frac{3}{4} & \frac{1}{4} \end{pmatrix}_i} \times \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i = \begin{pmatrix} \underbrace{\frac{1}{2} \times \frac{4}{3} \times \frac{1}{2}}_{1/3} + 0 \\ \underbrace{\frac{1}{2} \times \frac{4}{3} \times \frac{1}{4}}_{1/6} + \underbrace{\frac{1}{2} \times 4 \times \frac{1}{4}}_{1/2} \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \end{pmatrix}_j$$

After one iteration

⋮

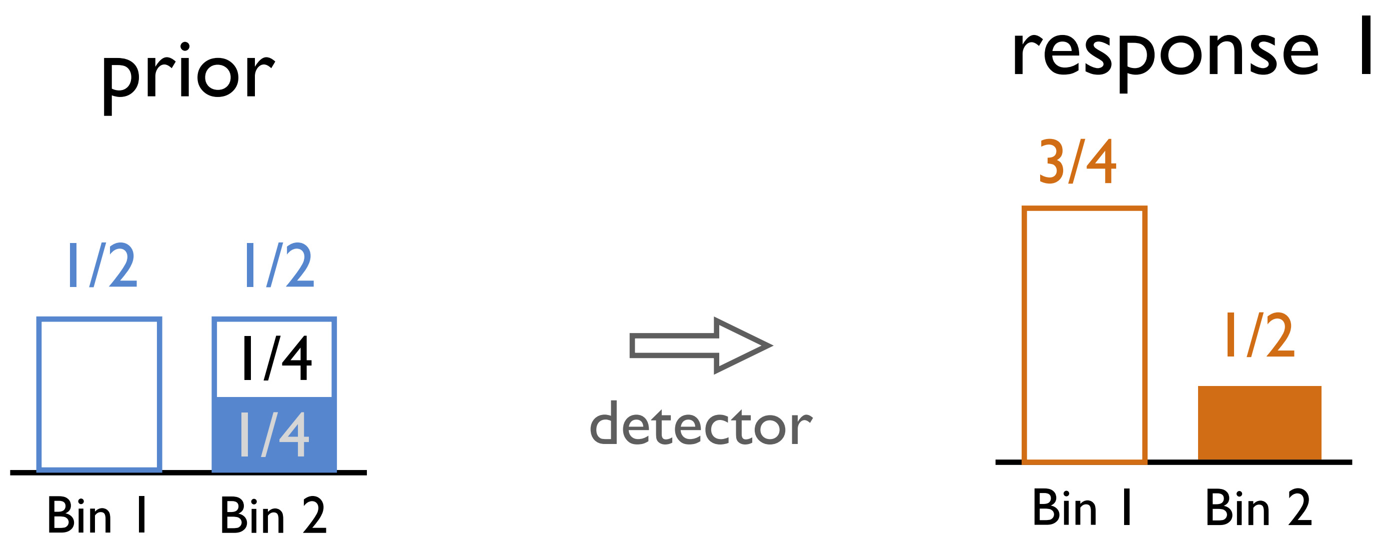
Correct truth distribution obtained as $n \to \infty$

$$t_j^{(n)} = \sum_i \frac{\begin{pmatrix} \frac{1}{n+1} & \frac{n}{2(n+1)} \\ 0 & \frac{n}{2(n+1)} \end{pmatrix}_{ij}}{\begin{pmatrix} \frac{n+2}{2(n+1)} & \frac{n}{2(n+1)} \end{pmatrix}_i} \times \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i = \begin{pmatrix} \frac{1}{n+2} \\ \frac{n+1}{n+2} \end{pmatrix}_j \to \begin{pmatrix} 0 \\ 1 \end{pmatrix}_j$$

At the n[th] iteration

IBU as Reweighting

[Richardson, JOSA 1972; Lucy, AJ 1974; D'Agostini, NIMPA 1995]



Consider a situation with two particle-level bins and two detector-level bins

$$t_j^{(0)} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_j \qquad m_i = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i \qquad R_{ij} = \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{pmatrix}_{ij}$$

Uniform prior

Bins are measured equally

Bin 1 reconstructed perfectly
Bin 2 reconstructed equally

$$t_j^{(1)} = \sum_i \frac{\begin{pmatrix} \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{4} \end{pmatrix}_{ij}}{\begin{pmatrix} \frac{3}{4} & \frac{1}{4} \end{pmatrix}_i} \times \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i = \begin{pmatrix} \underbrace{\frac{1}{2} \times \frac{4}{3} \times \frac{1}{2} + 0}_{1/3} \\ \underbrace{\frac{1}{2} \times \frac{4}{3} \times \frac{1}{4}}_{1/6} + \underbrace{\frac{1}{2} \times 4 \times \frac{1}{4}}_{1/2} \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \end{pmatrix}_j$$

After one iteration

$\vdots$

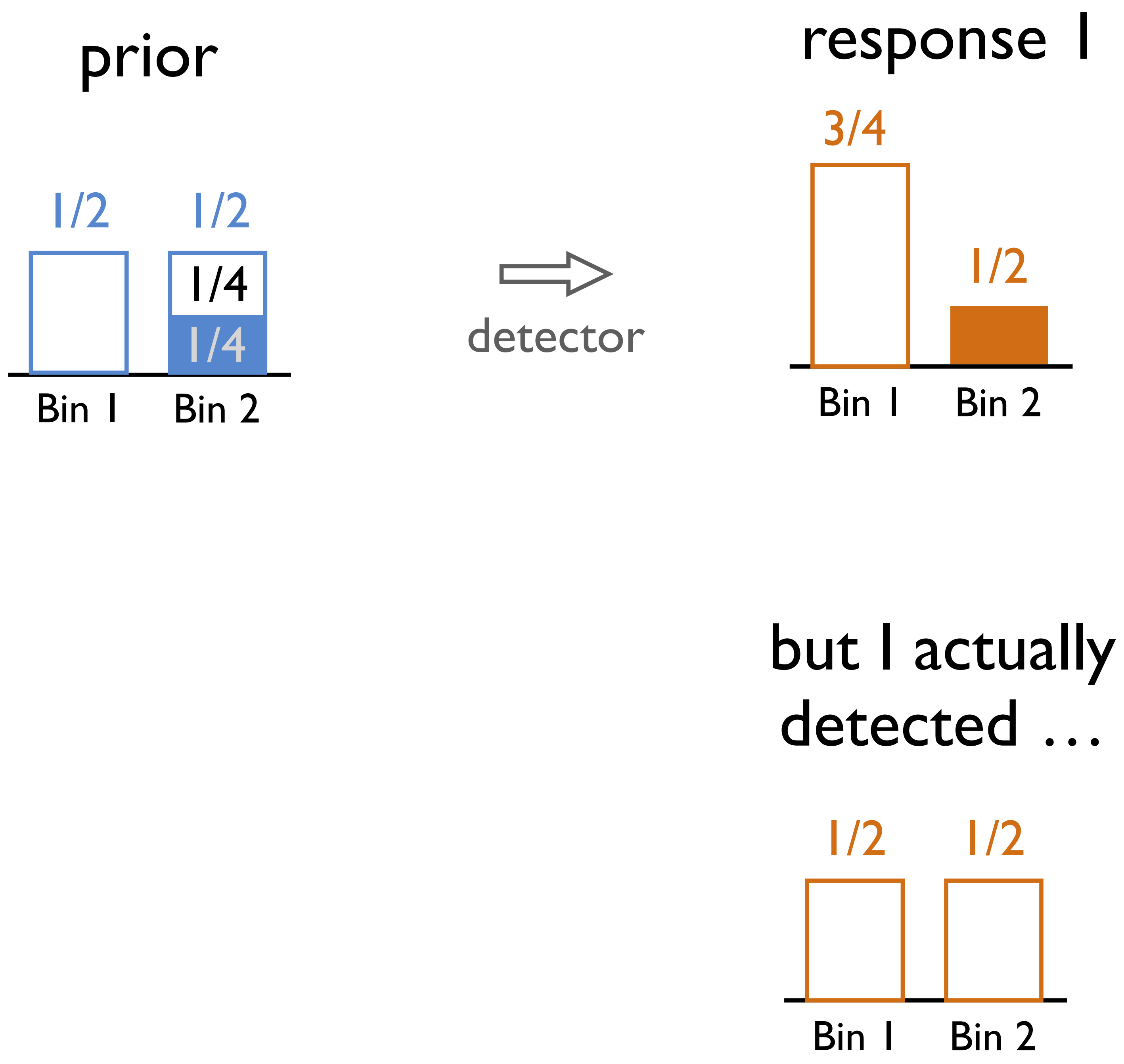
Correct truth distribution obtained as $n \to \infty$

$$t_j^{(n)} = \sum_i \frac{\begin{pmatrix} \frac{1}{n+1} & \frac{n}{2(n+1)} \\ 0 & \frac{n}{2(n+1)} \end{pmatrix}_{ij}}{\begin{pmatrix} \frac{n+2}{2(n+1)} & \frac{n}{2(n+1)} \end{pmatrix}_i} \times \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i = \begin{pmatrix} \frac{1}{n+2} \\ \frac{n+1}{n+2} \end{pmatrix}_j \to \begin{pmatrix} 0 \\ 1 \end{pmatrix}_j$$

At the n^{th} iteration

IBU as Reweighting

[Richardson, JOSA 1972; Lucy, AJ 1974; D'Agostini, NIMPA 1995]



Consider a situation with two particle-level bins and two detector-level bins

$$t_j^{(0)} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_j \qquad m_i = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i \qquad R_{ij} = \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{pmatrix}_{ij}$$

Uniform prior

Bins are measured equally

Bin 1 reconstructed perfectly
Bin 2 reconstructed equally

$$t_j^{(1)} = \sum_i \frac{\begin{pmatrix} \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{4} \end{pmatrix}_{ij}}{\begin{pmatrix} \frac{3}{4} & \frac{1}{4} \end{pmatrix}_i} \times \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i = \begin{pmatrix} \underbrace{\frac{1}{2} \times \frac{4}{3} \times \frac{1}{2} + 0}_{1/3} \\ \underbrace{\frac{1}{2} \times \frac{4}{3} \times \frac{1}{4}}_{1/6} + \underbrace{\frac{1}{2} \times 4 \times \frac{1}{4}}_{1/2} \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \end{pmatrix}_j$$

After one iteration

⋮

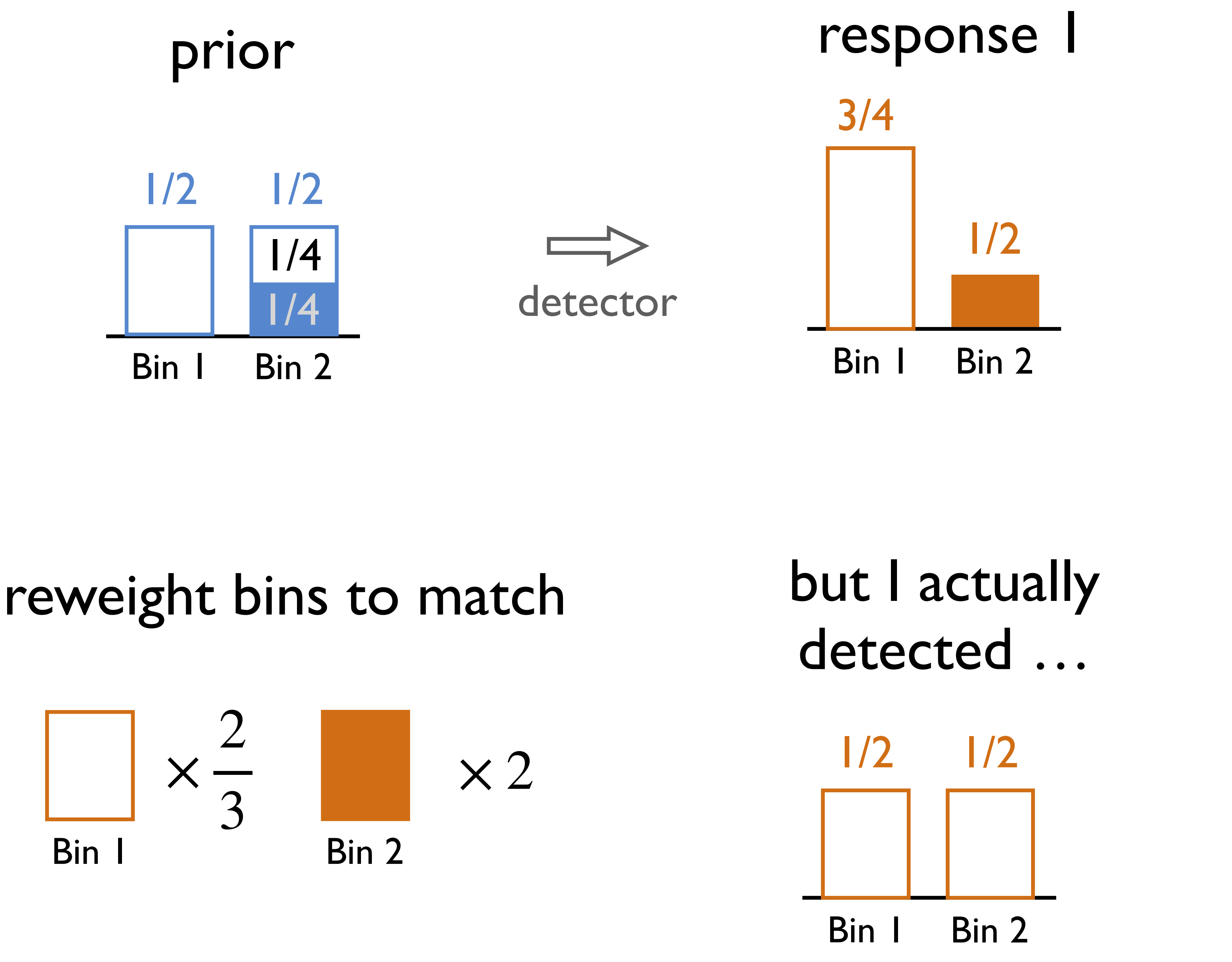
Correct truth distribution obtained as $n \to \infty$

$$t_j^{(n)} = \sum_i \frac{\begin{pmatrix} \frac{1}{n+1} & \frac{n}{2(n+1)} \\ 0 & \frac{n}{2(n+1)} \end{pmatrix}_{ij}}{\begin{pmatrix} \frac{n+2}{2(n+1)} & \frac{n}{2(n+1)} \end{pmatrix}_i} \times \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i = \begin{pmatrix} \frac{1}{n+2} \\ \frac{n+1}{n+2} \end{pmatrix}_j \to \begin{pmatrix} 0 \\ 1 \end{pmatrix}_j$$

At the nth iteration

IBU as Reweighting

[Richardson, JOSA 1972; Lucy, AJ 1974; D'Agostini, NIMPA 1995]



Consider a situation with two particle-level bins and two detector-level bins

$$t_j^{(0)} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_j \qquad m_i = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i \qquad R_{ij} = \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{pmatrix}_{ij}$$

Uniform prior

Bins are measured equally

Bin 1 reconstructed perfectly
Bin 2 reconstructed equally

$$t_j^{(1)} = \sum_i \frac{\begin{pmatrix} \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{4} \end{pmatrix}_{ij}}{\begin{pmatrix} \frac{3}{4} & \frac{1}{4} \end{pmatrix}_i} \times \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i = \begin{pmatrix} \underbrace{\frac{1}{2} \times \frac{4}{3} \times \frac{1}{2}}_{1/3} + 0 \\ \underbrace{\frac{1}{2} \times \frac{4}{3} \times \frac{1}{4}}_{1/6} + \underbrace{\frac{1}{2} \times 4 \times \frac{1}{4}}_{1/2} \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \end{pmatrix}_j$$

After one iteration

$\vdots$

Correct truth distribution obtained as $n \to \infty$

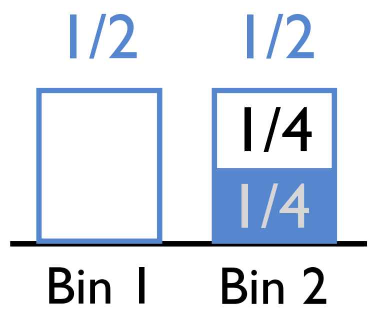
$$t_j^{(n)} = \sum_i \frac{\begin{pmatrix} \frac{1}{n+1} & \frac{n}{2(n+1)} \\ 0 & \frac{n}{2(n+1)} \end{pmatrix}_{ij}}{\begin{pmatrix} \frac{n+2}{2(n+1)} & \frac{n}{2(n+1)} \end{pmatrix}_i} \times \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i = \begin{pmatrix} \frac{1}{n+2} \\ \frac{n+1}{n+2} \end{pmatrix}_j \to \begin{pmatrix} 0 \\ 1 \end{pmatrix}_j$$

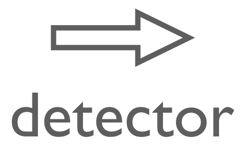
At the n^{th} iteration

IBU as Reweighting

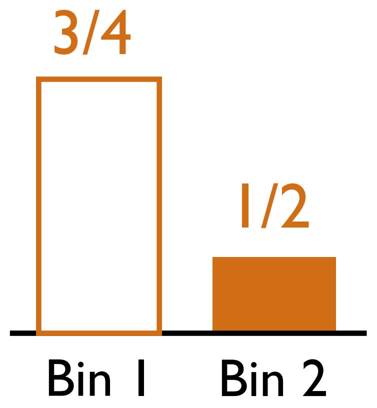
[Richardson, JOSA 1972; Lucy, AJ 1974; D'Agostini, NIMPA 1995]

prior

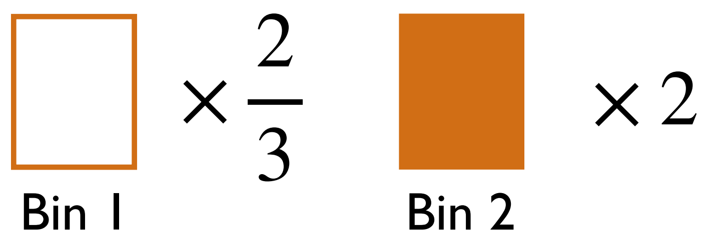




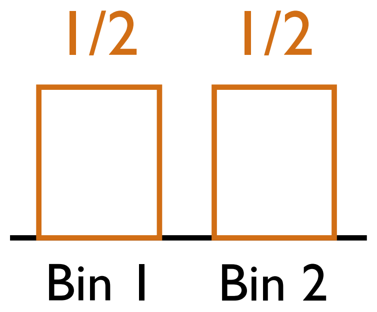
response 1



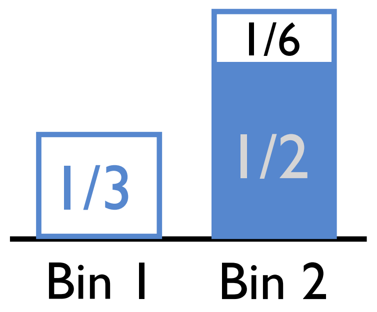
reweight bins to match



but I actually detected …



pull reweighting back to truth level



Consider a situation with two particle-level bins and two detector-level bins

$$t_j^{(0)} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_j \qquad m_i = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i \qquad R_{ij} = \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{pmatrix}_{ij}$$

Uniform prior

Bins are measured equally

Bin 1 reconstructed perfectly
Bin 2 reconstructed equally

$$t_j^{(1)} = \sum_i \frac{\begin{pmatrix} \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{4} \end{pmatrix}_{ij}}{\begin{pmatrix} \frac{3}{4} & \frac{1}{4} \end{pmatrix}_i} \times \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i = \begin{pmatrix} \underbrace{\frac{1}{2} \times \frac{4}{3} \times \frac{1}{2}}_{1/3} + 0 \\ \underbrace{\frac{1}{2} \times \frac{4}{3} \times \frac{1}{4}}_{1/6} + \underbrace{\frac{1}{2} \times 4 \times \frac{1}{4}}_{1/2} \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \end{pmatrix}_j$$

After one iteration

⋮

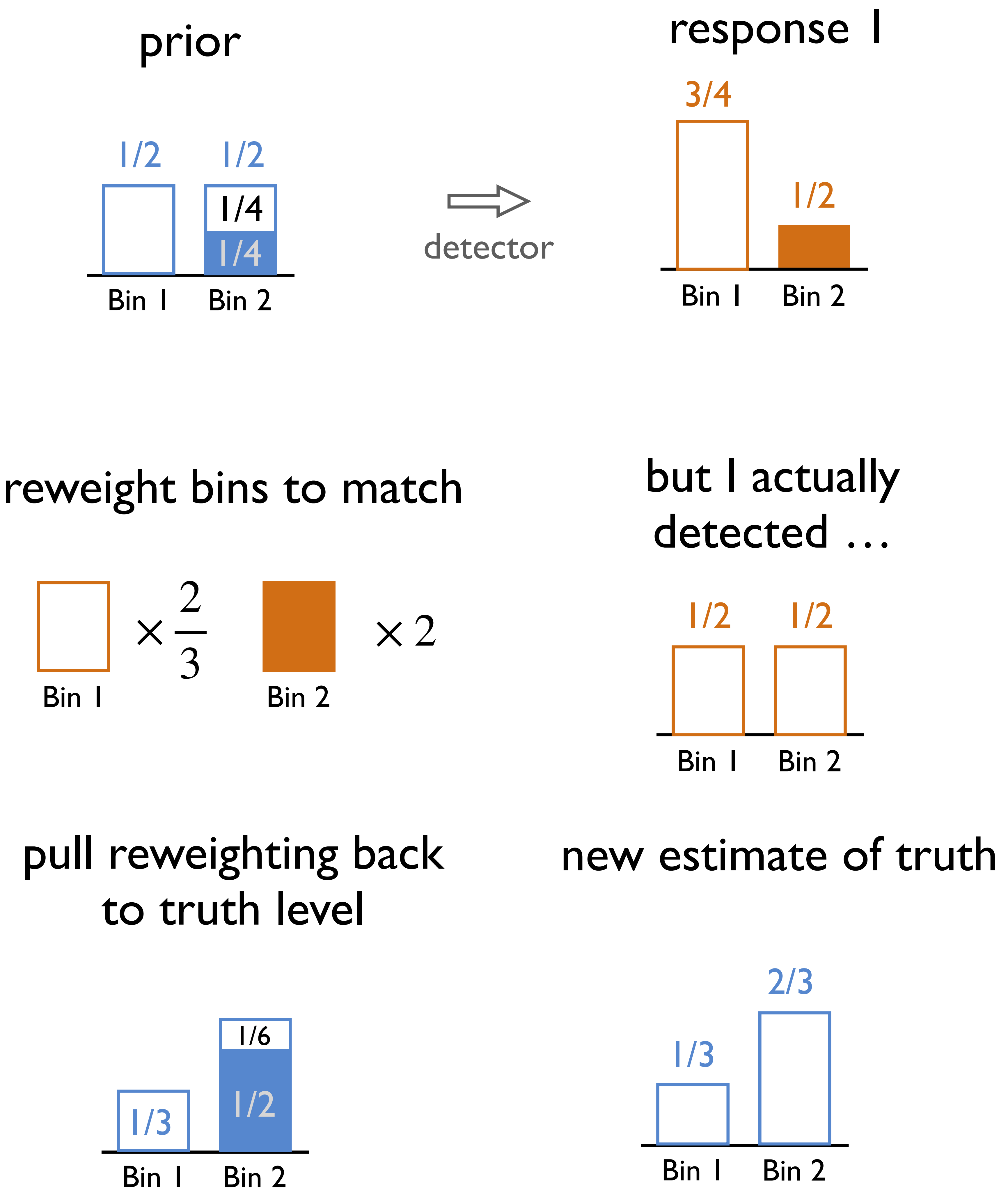
$$t_j^{(n)} = \sum_i \frac{\begin{pmatrix} \frac{1}{n+1} & \frac{n}{2(n+1)} \\ 0 & \frac{n}{2(n+1)} \end{pmatrix}_{ij}}{\begin{pmatrix} \frac{n+2}{2(n+1)} & \frac{n}{2(n+1)} \end{pmatrix}_i} \times \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i = \begin{pmatrix} \frac{1}{n+2} \\ \frac{n+1}{n+2} \end{pmatrix}_j \to \begin{pmatrix} 0 \\ 1 \end{pmatrix}_j$$

At the n^{th} iteration

Correct truth distribution obtained as $n \to \infty$

IBU as Reweighting

[Richardson, JOSA 1972; Lucy, AJ 1974; D'Agostini, NIMPA 1995]



Consider a situation with two particle-level bins and two detector-level bins

$$t_j^{(0)} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_j \qquad m_i = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i \qquad R_{ij} = \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{pmatrix}_{ij}$$

Uniform prior — Bins are measured equally — Bin 1 reconstructed perfectly, Bin 2 reconstructed equally

$$t_j^{(1)} = \sum_i \frac{\begin{pmatrix} \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{4} \end{pmatrix}_{ij}}{\begin{pmatrix} \frac{3}{4} & \frac{1}{4} \end{pmatrix}_i} \times \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i = \begin{pmatrix} \underbrace{\frac{1}{2} \times \frac{4}{3} \times \frac{1}{2}}_{1/3} + 0 \\ \underbrace{\frac{1}{2} \times \frac{4}{3} \times \frac{1}{4}}_{1/6} + \underbrace{\frac{1}{2} \times 4 \times \frac{1}{4}}_{1/2} \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \end{pmatrix}_j$$

After one iteration

⋮

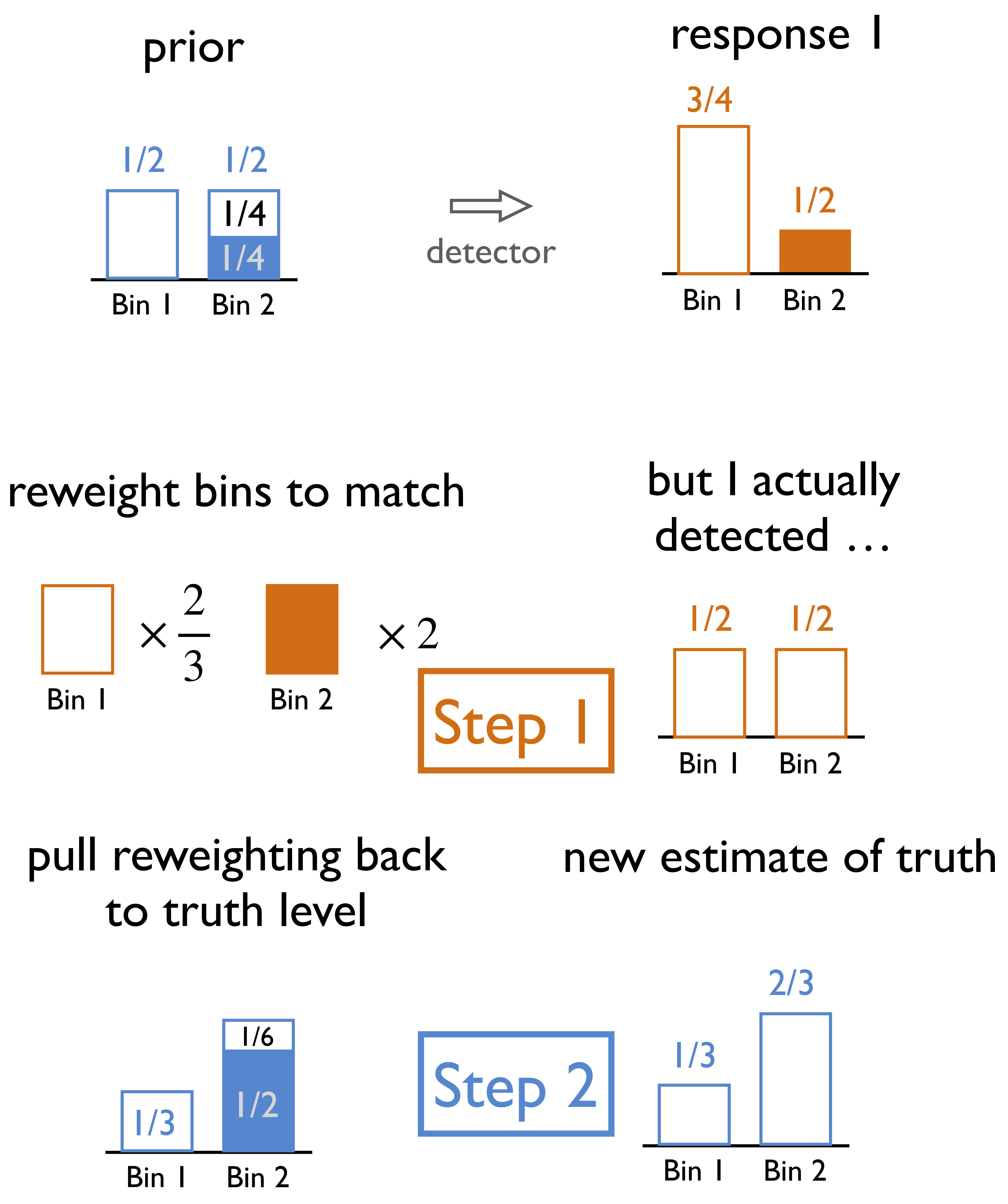
$$t_j^{(n)} = \sum_i \frac{\begin{pmatrix} \frac{1}{n+1} & \frac{n}{2(n+1)} \\ 0 & \frac{n}{2(n+1)} \end{pmatrix}_{ij}}{\begin{pmatrix} \frac{n+2}{2(n+1)} & \frac{n}{2(n+1)} \end{pmatrix}_i} \times \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i = \begin{pmatrix} \frac{1}{n+2} \\ \frac{n+1}{n+2} \end{pmatrix}_j \to \begin{pmatrix} 0 \\ 1 \end{pmatrix}_j$$

Correct truth distribution obtained as $n \to \infty$

At the n[th] iteration

IBU as Reweighting

[Richardson, JOSA 1972; Lucy, AJ 1974; D'Agostini, NIMPA 1995]



Consider a situation with two particle-level bins and two detector-level bins

$$t_j^{(0)} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_j \qquad m_i = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i \qquad R_{ij} = \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{pmatrix}_{ij}$$

Uniform prior — Bins are measured equally — Bin 1 reconstructed perfectly, Bin 2 reconstructed equally

$$t_j^{(1)} = \sum_i \frac{\begin{pmatrix} \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{4} \end{pmatrix}_{ij}}{\begin{pmatrix} \frac{3}{4} & \frac{1}{4} \end{pmatrix}_i} \times \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i = \begin{pmatrix} \underbrace{\frac{1}{2} \times \frac{4}{3} \times \frac{1}{2}}_{1/3} + 0 \\ \underbrace{\frac{1}{2} \times \frac{4}{3} \times \frac{1}{4}}_{1/6} + \underbrace{\frac{1}{2} \times 4 \times \frac{1}{4}}_{1/2} \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \end{pmatrix}_j$$

After one iteration

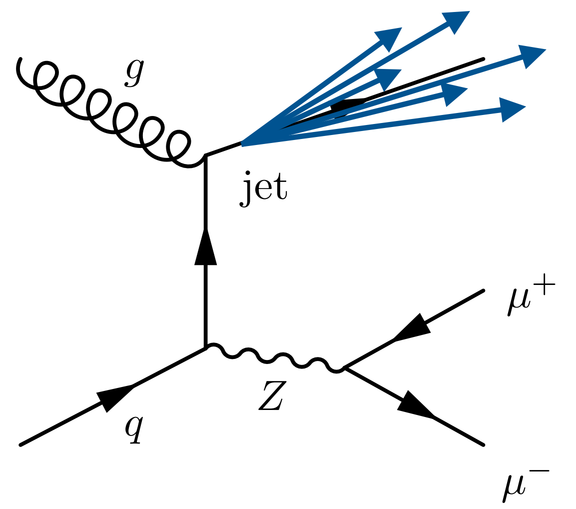
$\vdots$

$$t_j^{(n)} = \sum_i \frac{\begin{pmatrix} \frac{1}{n+1} & \frac{n}{2(n+1)} \\ 0 & \frac{n}{2(n+1)} \end{pmatrix}_{ij}}{\begin{pmatrix} \frac{n+2}{2(n+1)} & \frac{n}{2(n+1)} \end{pmatrix}_i} \times \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_i = \begin{pmatrix} \frac{1}{n+2} \\ \frac{n+1}{n+2} \end{pmatrix}_j \to \begin{pmatrix} 0 \\ 1 \end{pmatrix}_j$$

At the n^th iteration

Correct truth distribution obtained as $n \to \infty$

OmniFold Results by Event Representation



User is free to choose *event representation* in the OmniFold procedure

OMNIFOLD – full phase space information

↓

MULTIFOLD – multiple observables

↓

UNIFOLD – single observable, essentially unbinned IBU

Method	m	M	w	$\ln\rho$	τ_{21}	z_g
	Observable					
OMNIFOLD	**2.77**	**0.33**	0.10	**0.35**	0.53	0.68
MULTIFOLD	3.80	0.89	**0.09**	0.37	**0.26**	**0.15**
UNIFOLD	8.82	1.46	0.15	0.59	1.11	0.59
IBU	9.31	1.51	0.11	0.71	1.10	0.37
Data	24.6	130	15.7	14.2	11.1	3.76
Generation	3.62	15	22.4	19	20.8	3.84
	mass	mult.	width	groomed mass	N-subj. ratio	

Evaluate performance using triangular discriminator

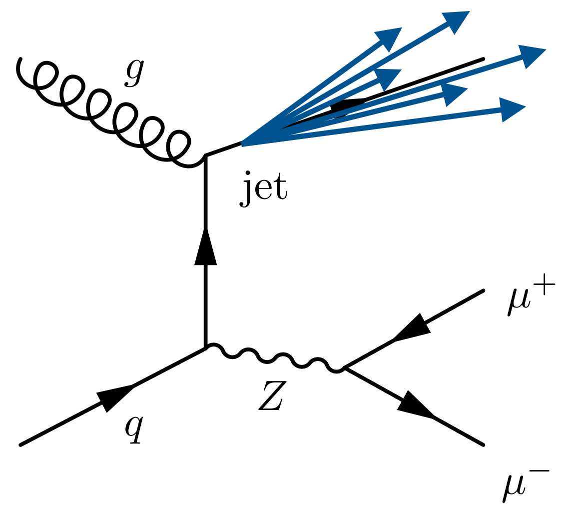
$$\Delta(p,q) = \frac{1}{2}\int d\lambda\, \frac{(p(\lambda)-q(\lambda))^2}{p(\lambda)+q(\lambda)} (\times 10^3)$$

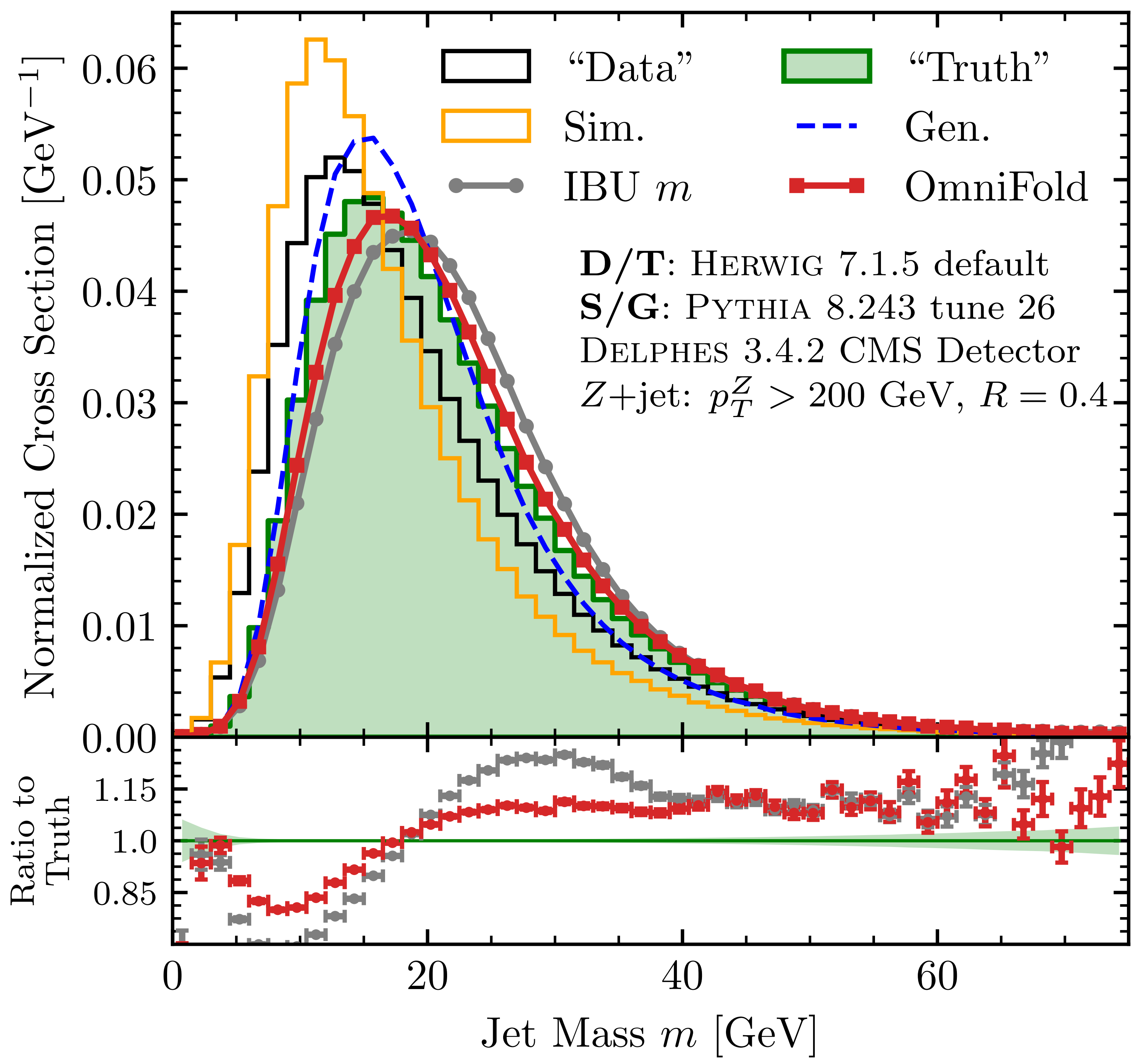
OMNIFOLD/MULTIFOLD *outperforms IBU on all observables!*

Single MULTIFOLD training based on all six observables

UNIFOLD is similar to or outperforms IBU

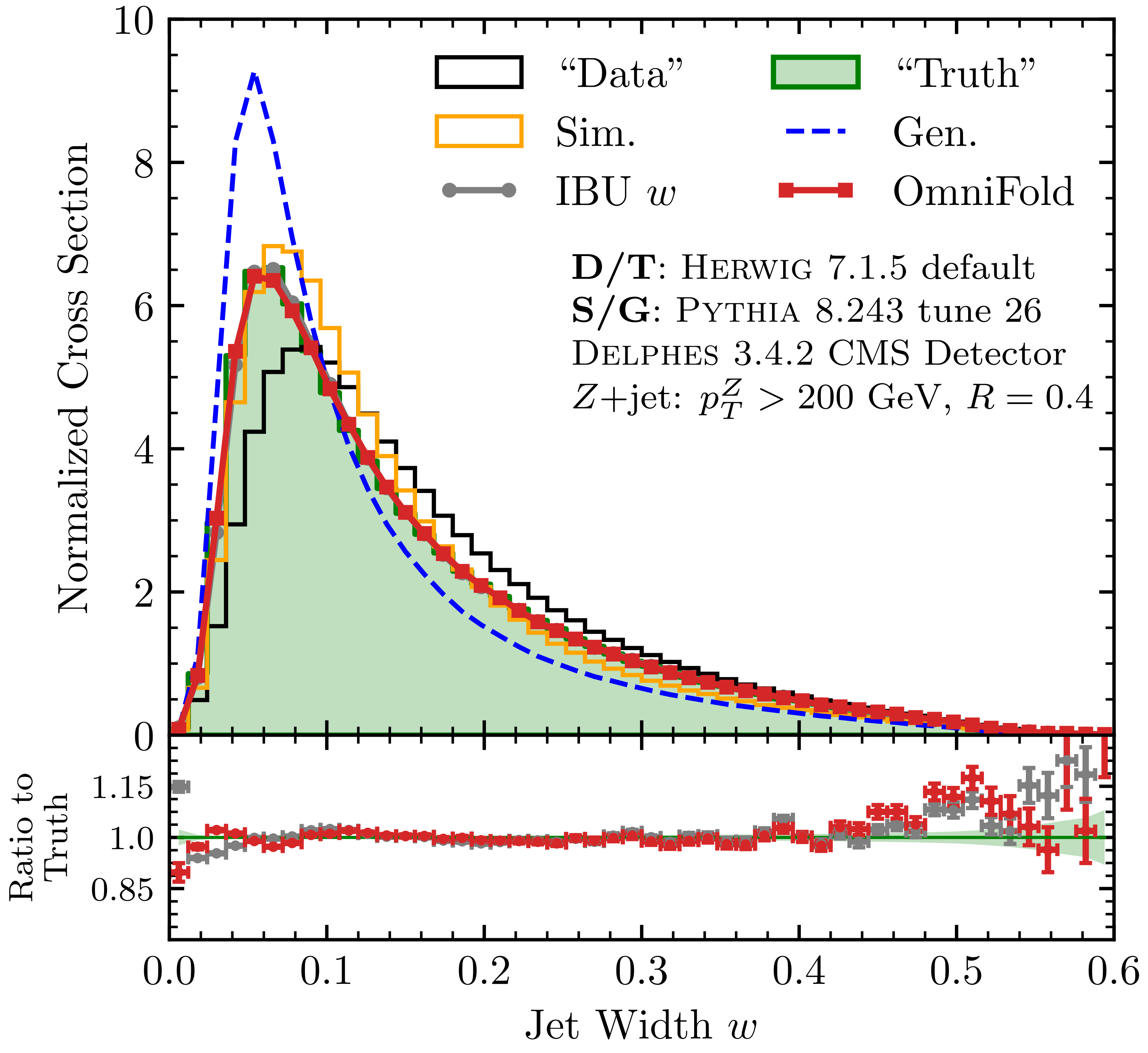
Additional OmniFolded Distributions





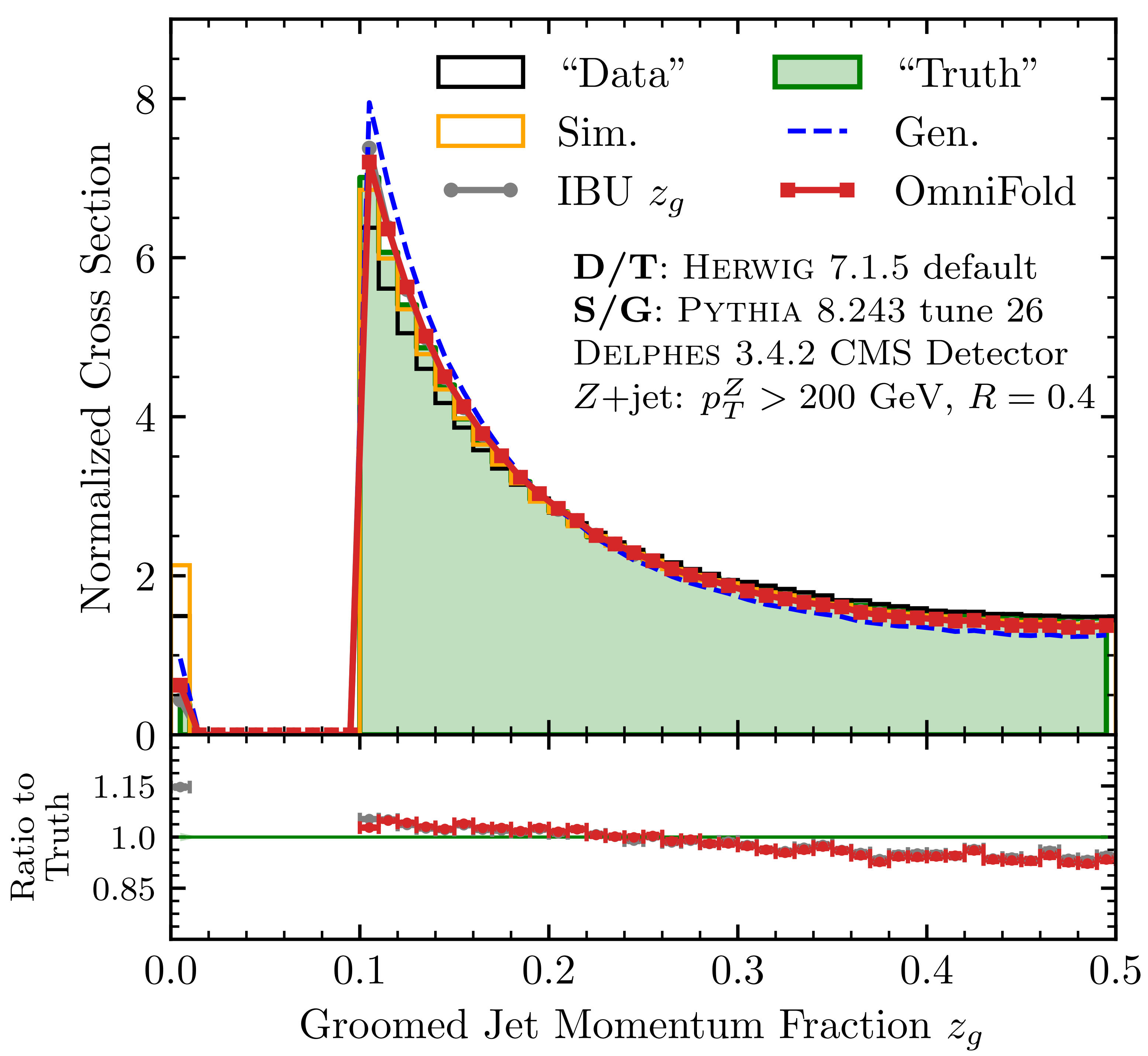
Jet mass affected by particle masses

$$m_J^2 = \left(\sum_{i \in \text{jet}} p_i^\mu \right)^2$$



IRC-safe observables easier to unfold

$$w = \frac{1}{\sum_j p_{Tj}} \sum_i p_{Ti} \sqrt{(y_i - y_J)^2 + (\phi_i - \phi_J)^2}$$



z_g remarkably stable under choice of method

z_g = pT fraction of first splitting to pass soft drop

OmniFold Etymology

The Mountain sat upon the Plain
In his tremendous Chair –
His observation **omnifold**,
His inquest, everywhere –

The Seasons played around his knees
Like Children round a sire –
Grandfather of the Days is He
Of Dawn, the Ancestor –

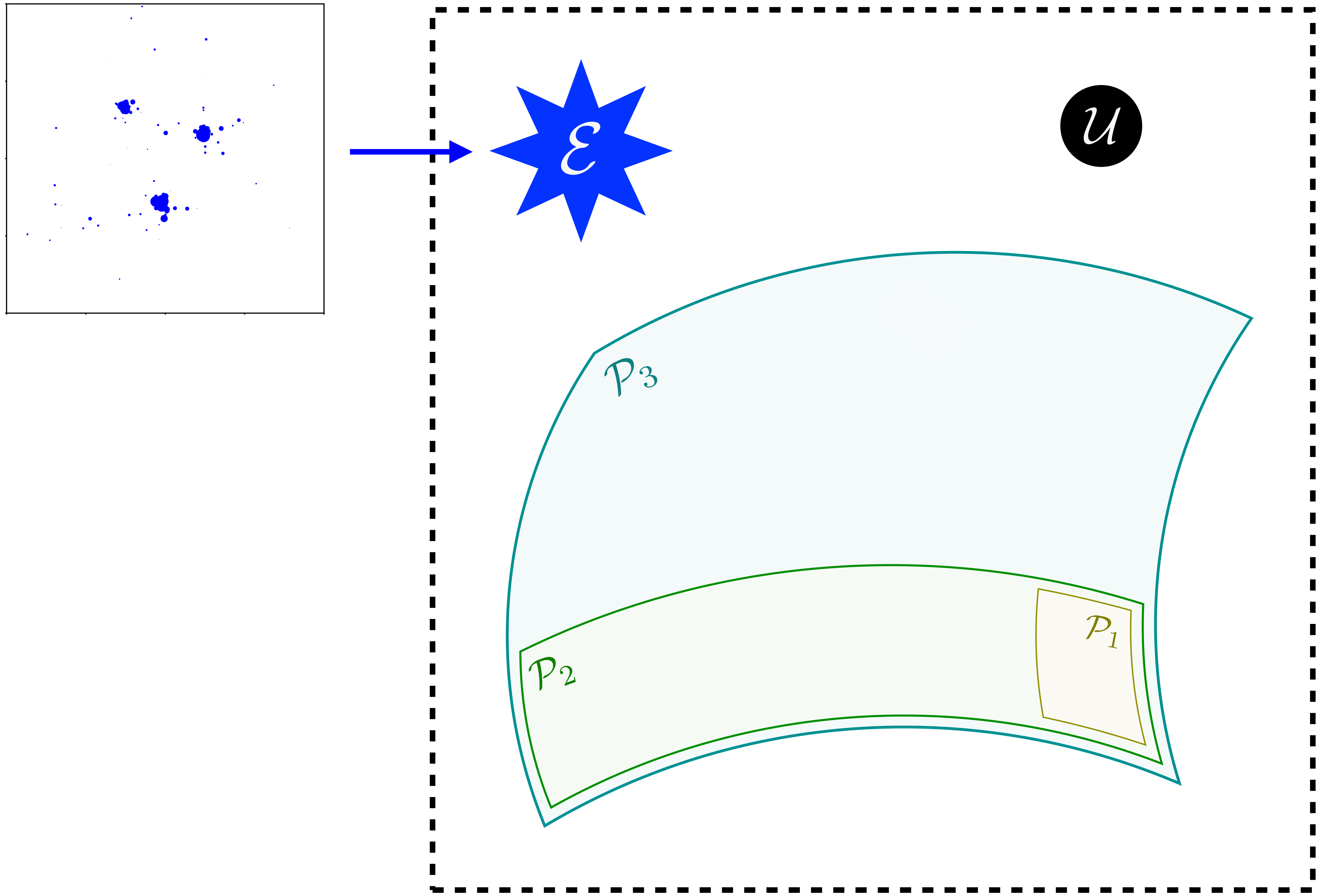
Emily Dickinson, #975

Energy Mover’s Distance

N-particle Manifolds in the Space of Events

[PTK, Metodiev, Thaler, JHEP 2020]

$\mathcal{P}_N$ = set of all N-particle configurations = $\left\{ \sum_{i=1}^{N} E_i\, \delta(\hat{n} - \hat{n}_i) \;\middle|\; E_i \geq 0 \right\}$



⋮

$\mathcal{P}_N \supset \mathcal{P}_{N-1} \supset \cdots \supset \mathcal{P}_3 \supset \mathcal{P}_2 \supset \mathcal{P}_1$

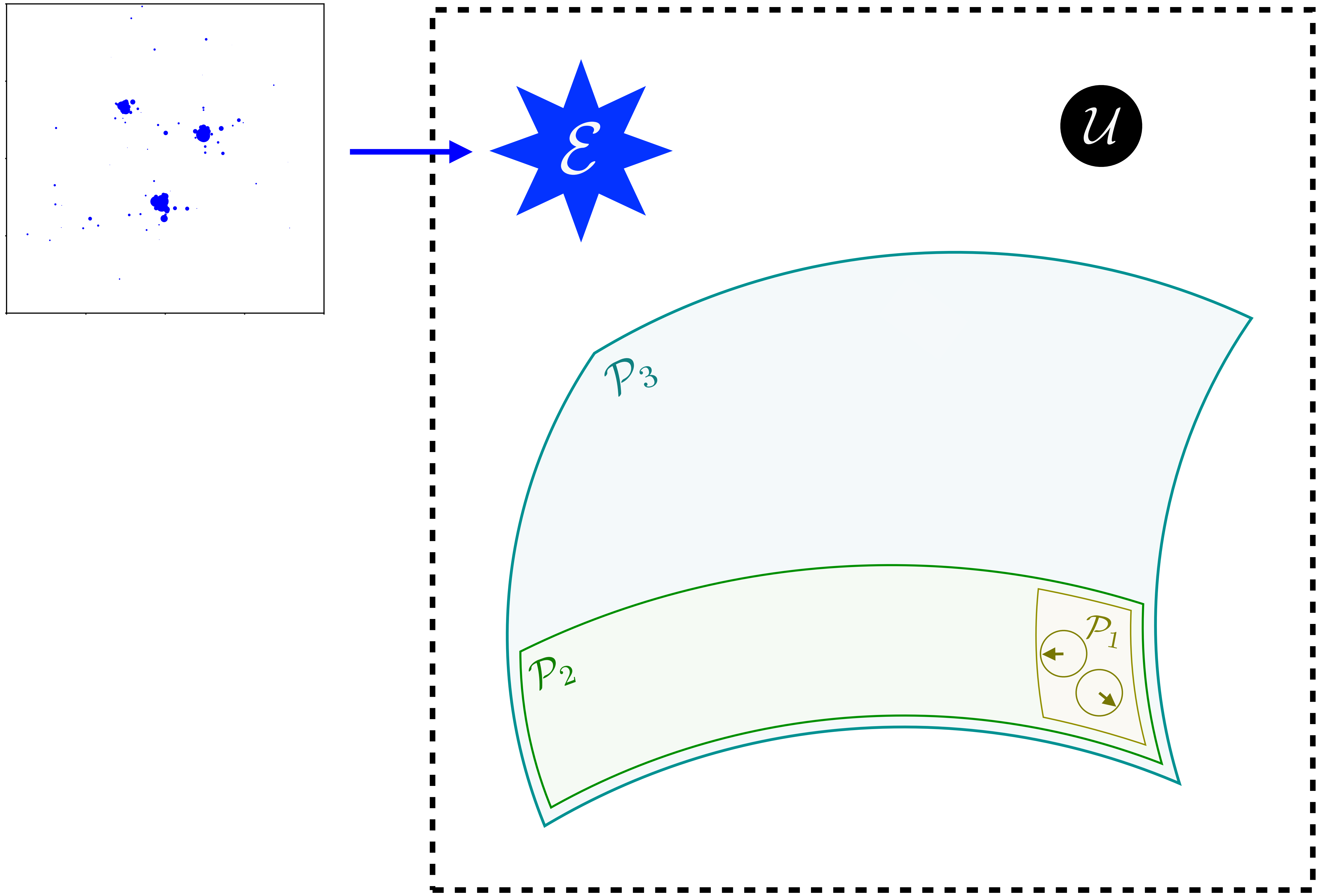
by soft and collinear limits

$\mathcal{U}$ Uniform event, not contained in any P_N

N-particle Manifolds in the Space of Events

[PTK, Metodiev, Thaler, JHEP 2020]

$\mathcal{P}_N$ = set of all N-particle configurations = $\left\{ \sum_{i=1}^{N} E_i\, \delta(\hat{n} - \hat{n}_i) \;\middle|\; E_i \geq 0 \right\}$



$\mathcal{P}_1$: manifold of events with one particle

⋮

$\mathcal{P}_N \supset \mathcal{P}_{N-1} \supset \cdots \supset \mathcal{P}_3 \supset \mathcal{P}_2 \supset \mathcal{P}_1$

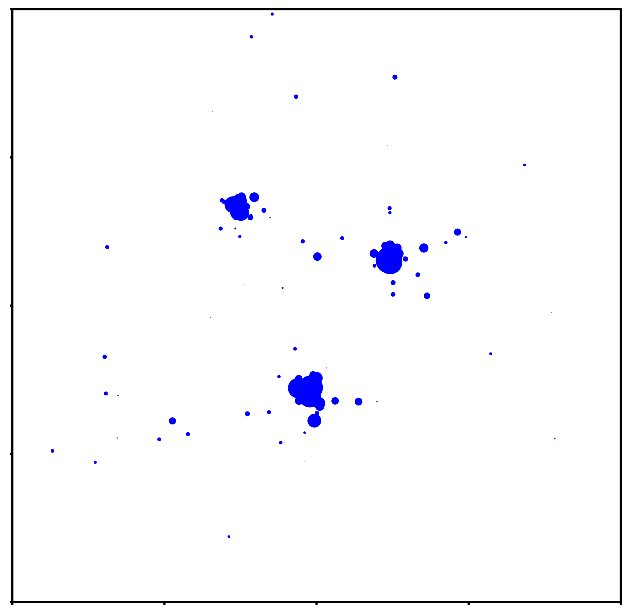
by soft and collinear limits

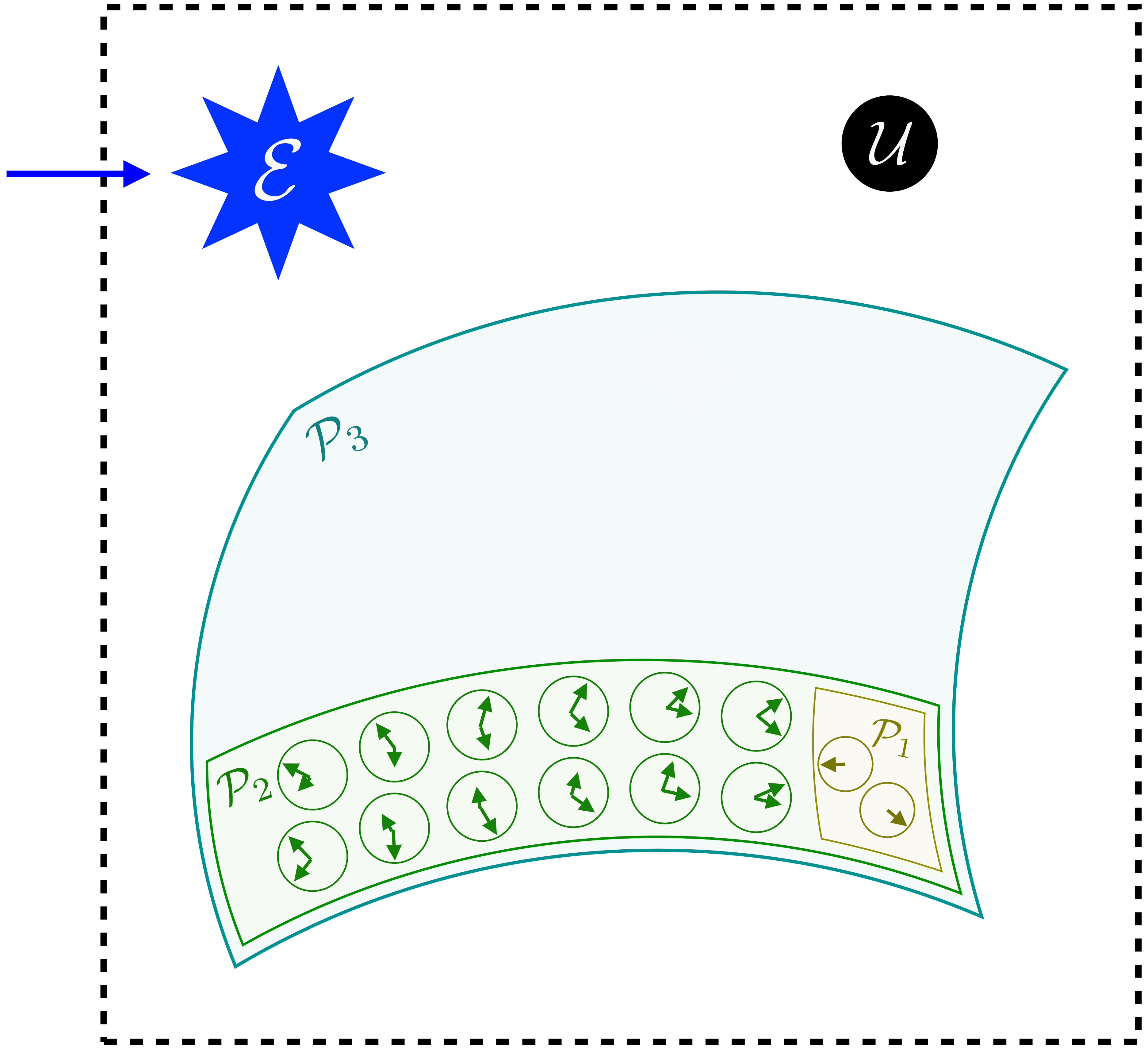
$\mathcal{U}$ Uniform event, not contained in any P_N

N-particle Manifolds in the Space of Events

[PTK, Metodiev, Thaler, JHEP 2020]

$\mathcal{P}_N$ = set of all N-particle configurations = $\left\{ \sum_{i=1}^{N} E_i\, \delta(\hat{n} - \hat{n}_i) \;\middle|\; E_i \geq 0 \right\}$





$\mathcal{P}_1$: manifold of events with one particle

$\mathcal{P}_2$: manifold of events with two particles

$\vdots$

$\mathcal{P}_N \supset \mathcal{P}_{N-1} \supset \cdots \supset \mathcal{P}_3 \supset \mathcal{P}_2 \supset \mathcal{P}_1$

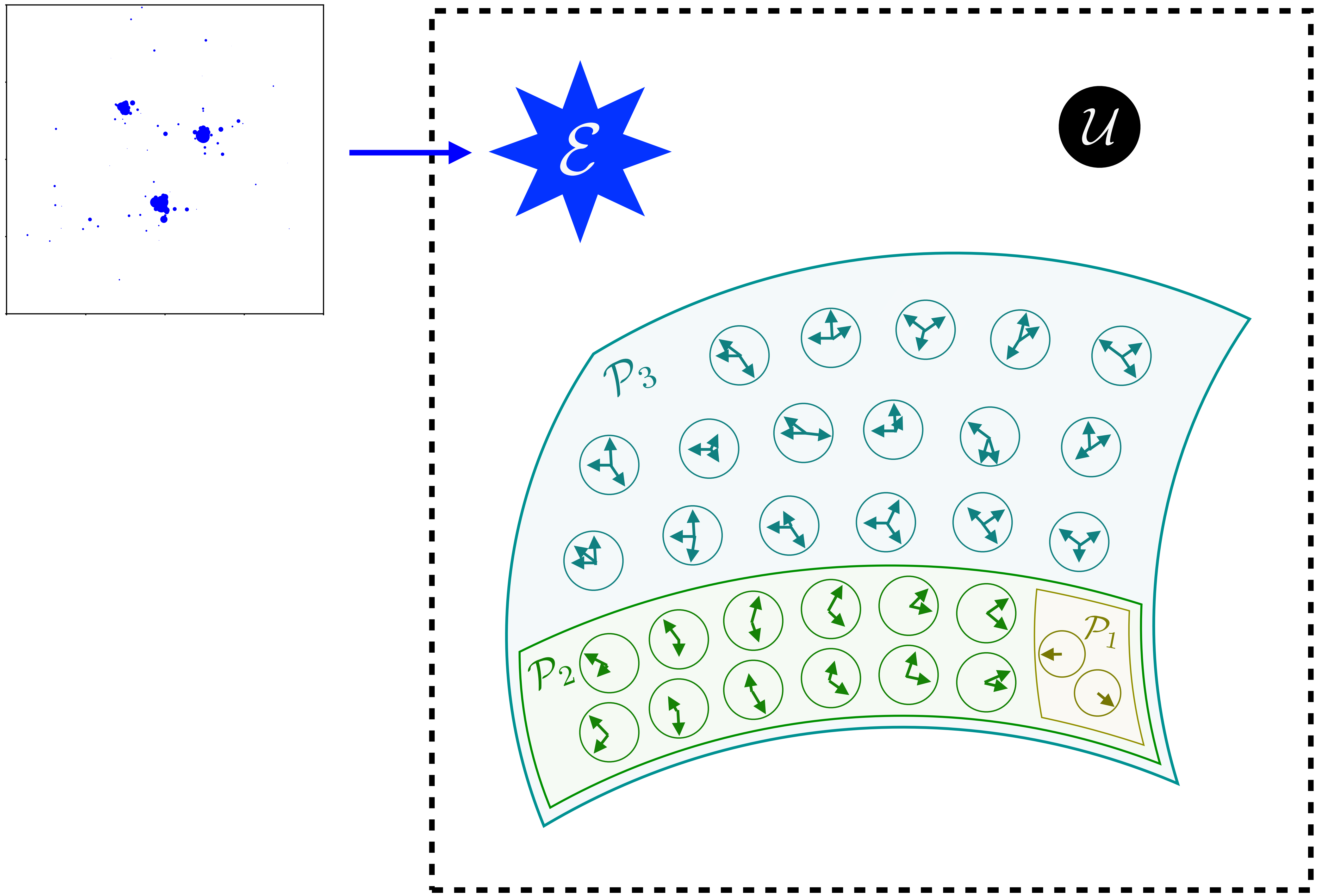
by soft and collinear limits

$\mathcal{U}$ Uniform event, not contained in any P_N

N-particle Manifolds in the Space of Events

[PTK, Metodiev, Thaler, JHEP 2020]

$\mathcal{P}_N$ = set of all N-particle configurations = $\left\{ \sum_{i=1}^{N} E_i\, \delta(\hat{n} - \hat{n}_i) \;\middle|\; E_i \geq 0 \right\}$



$\mathcal{P}_1$: manifold of events with one particle

$\mathcal{P}_2$: manifold of events with two particles

$\mathcal{P}_3$: manifold of events with three particles

⋮

$$\mathcal{P}_N \supset \mathcal{P}_{N-1} \supset \cdots \supset \mathcal{P}_3 \supset \mathcal{P}_2 \supset \mathcal{P}_1$$

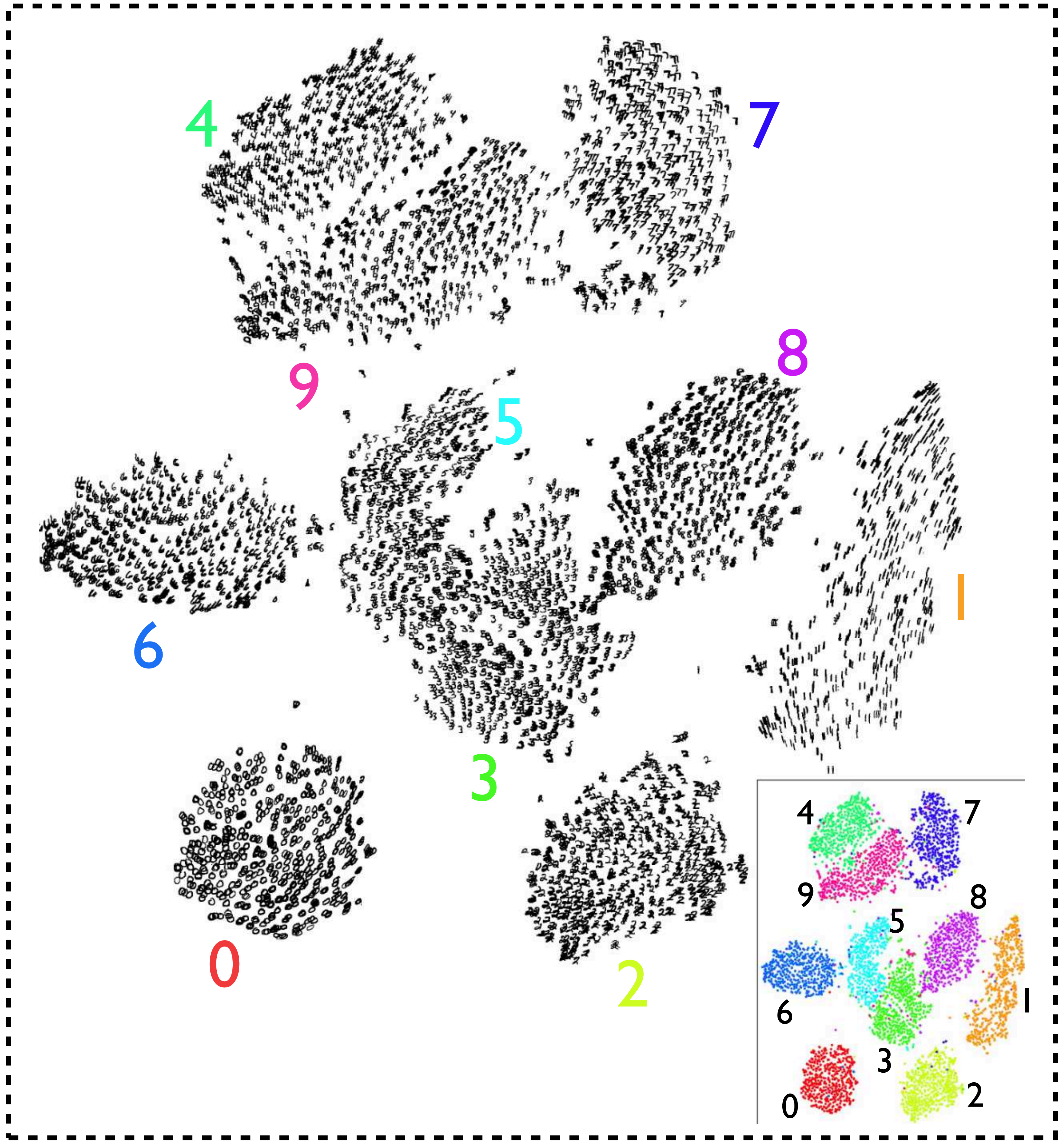
by soft and collinear limits

$\mathcal{U}$ Uniform event, not contained in any P_N

Visualizing Geometry in the Space of Events

t-Distributed Stochastic Neighbor Embedding (t-SNE)

MNIST handwritten digits

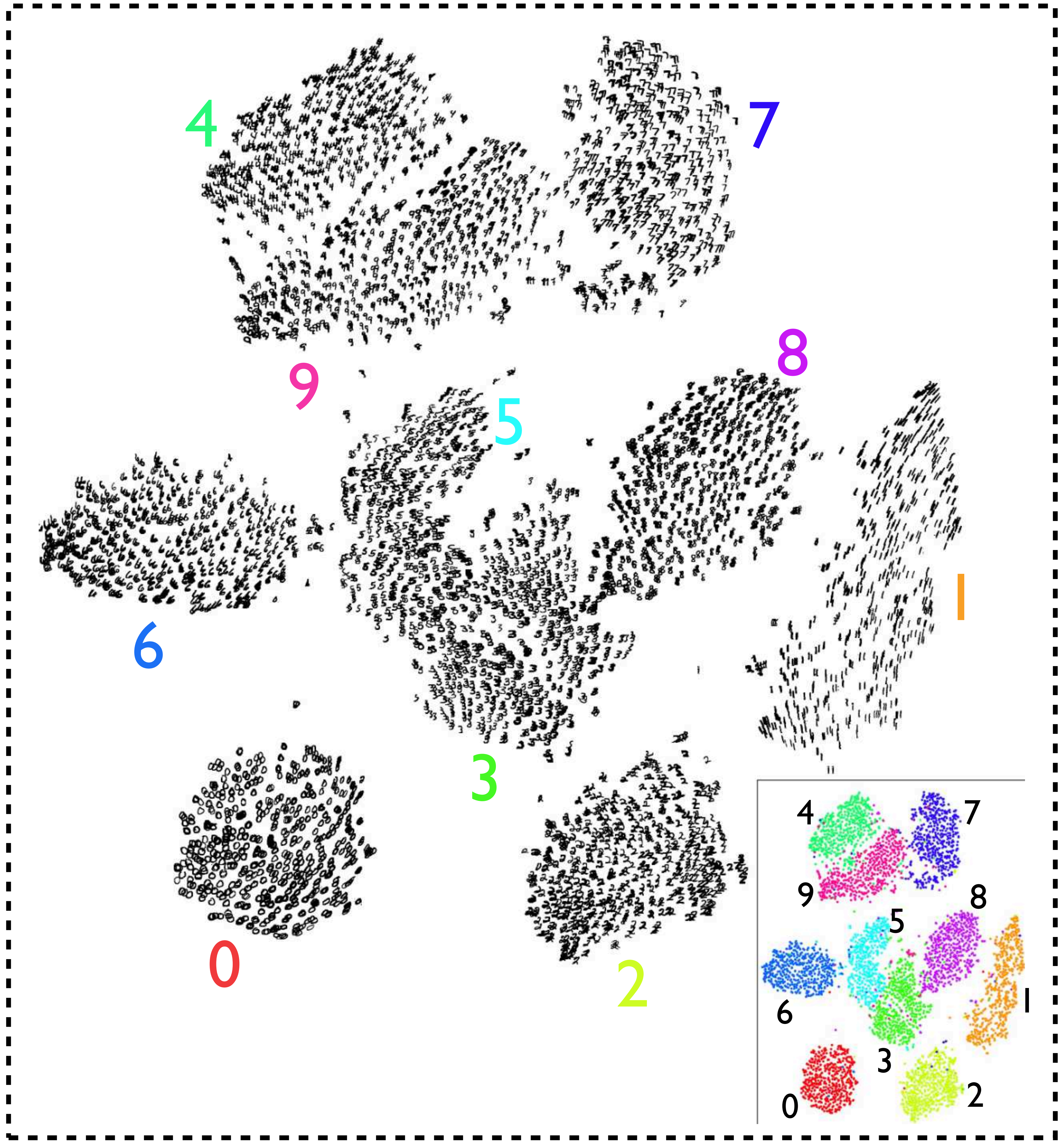


[L. van der Maaten, G. Hinton, JMLR 2008]

Visualizing Geometry in the Space of Events

t-Distributed Stochastic Neighbor Embedding (t-SNE)

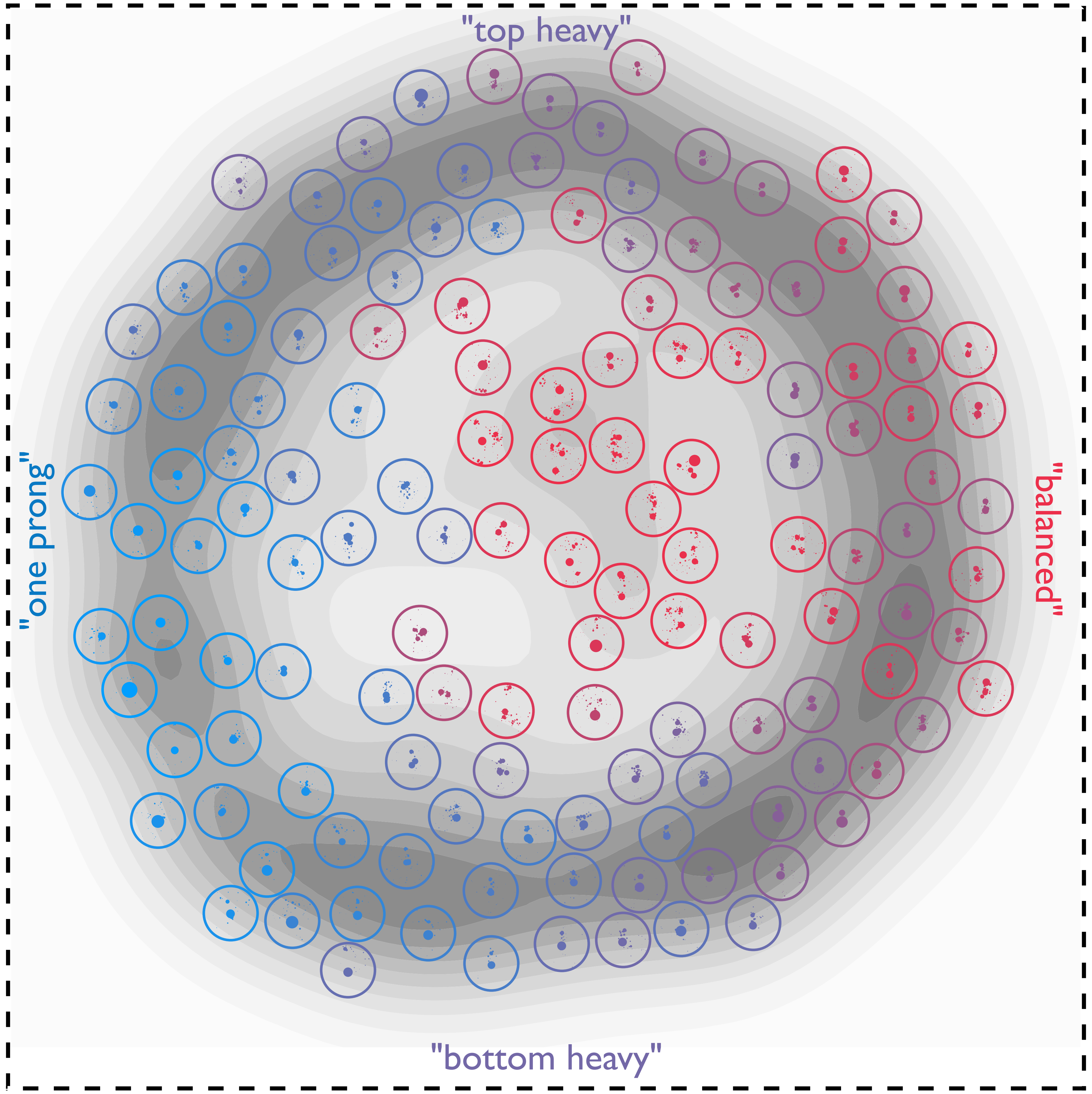
MNIST handwritten digits



[L. van der Maaten, G. Hinton, JMLR 2008]

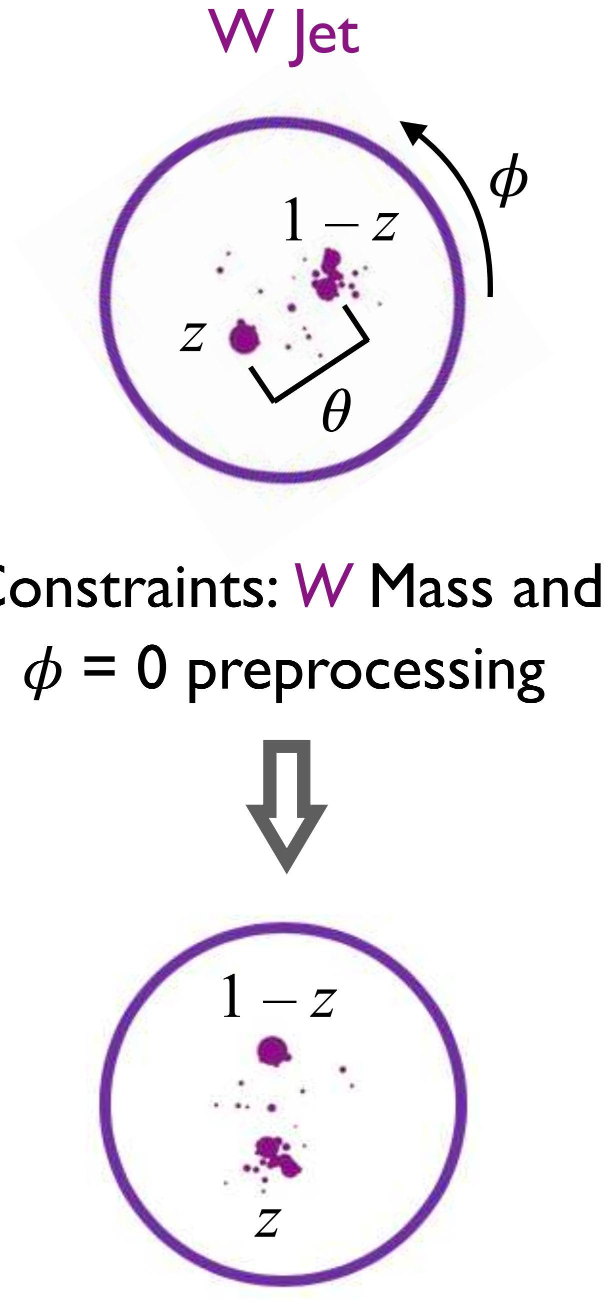
[PTK, Metodiev, Thaler, PRL 2019]

Geometric space of W jets



Gray contours represent the density of jets

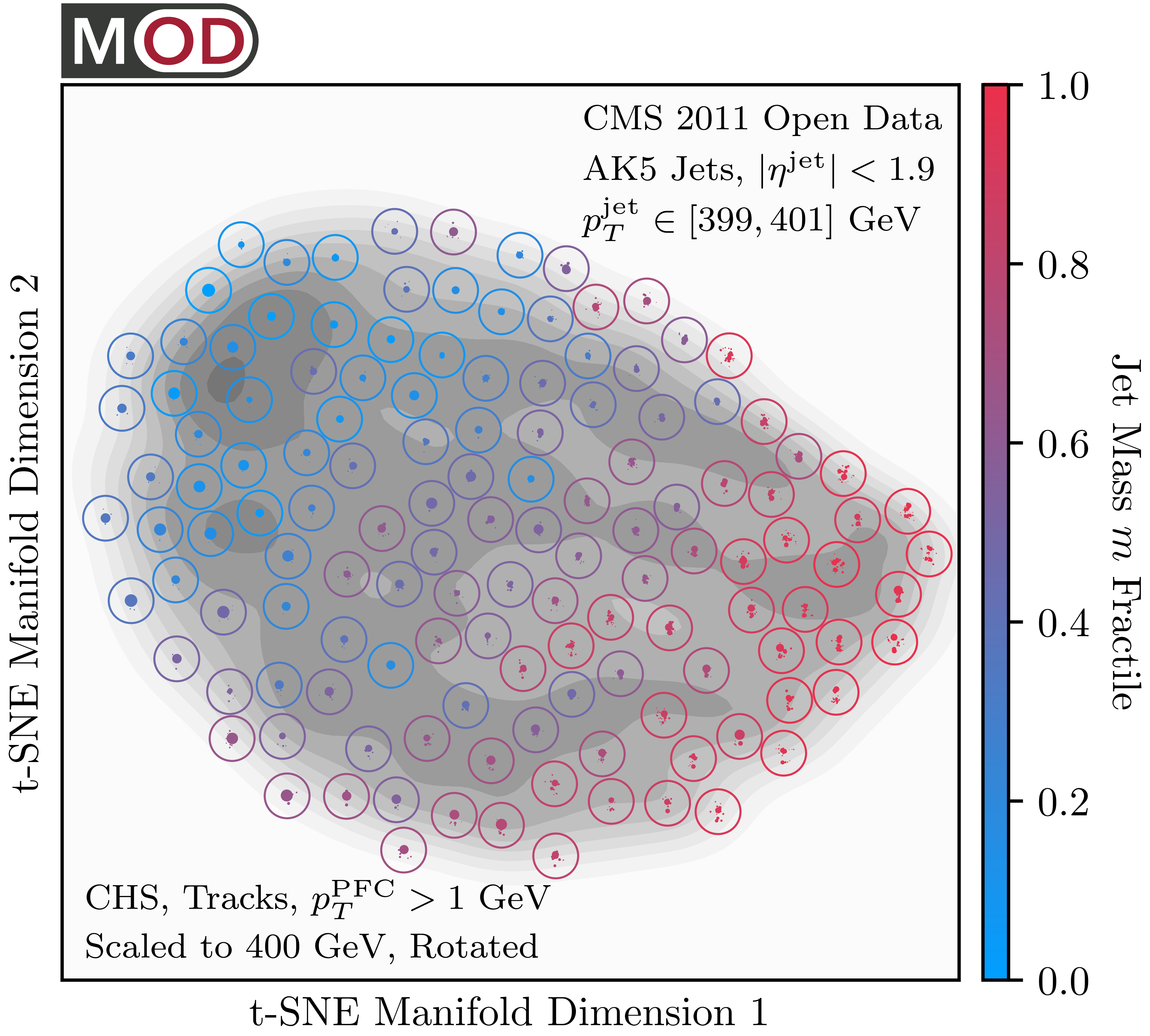
Each circle is a particular *W* jet



Constraints: *W* Mass and ϕ = 0 preprocessing

Visualizing Geometry in CMS Open Data

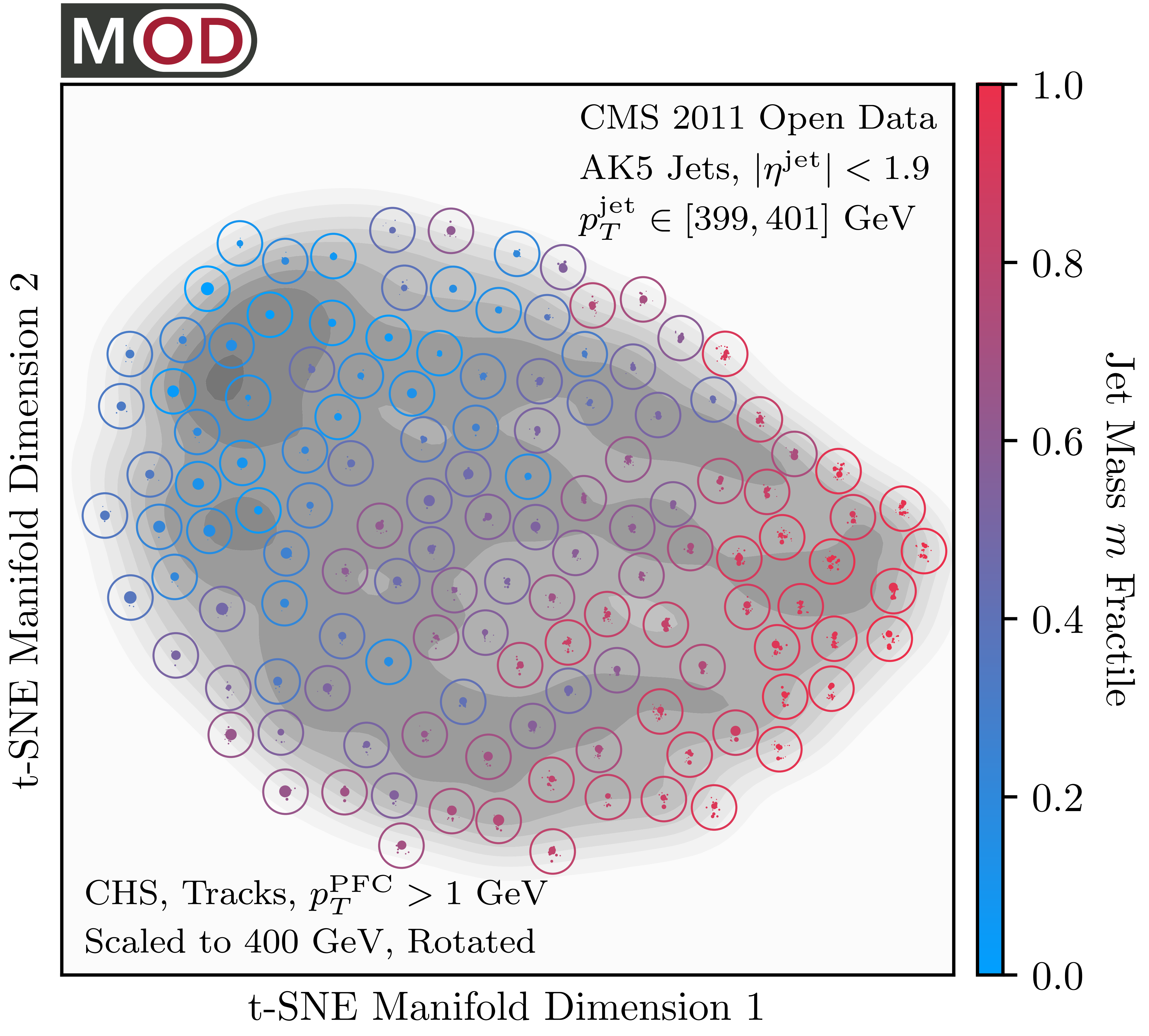
[PTK, Mastandrea, Metodiev, Naik, Thaler, PRD 2019; code and datasets at energyflow.network]



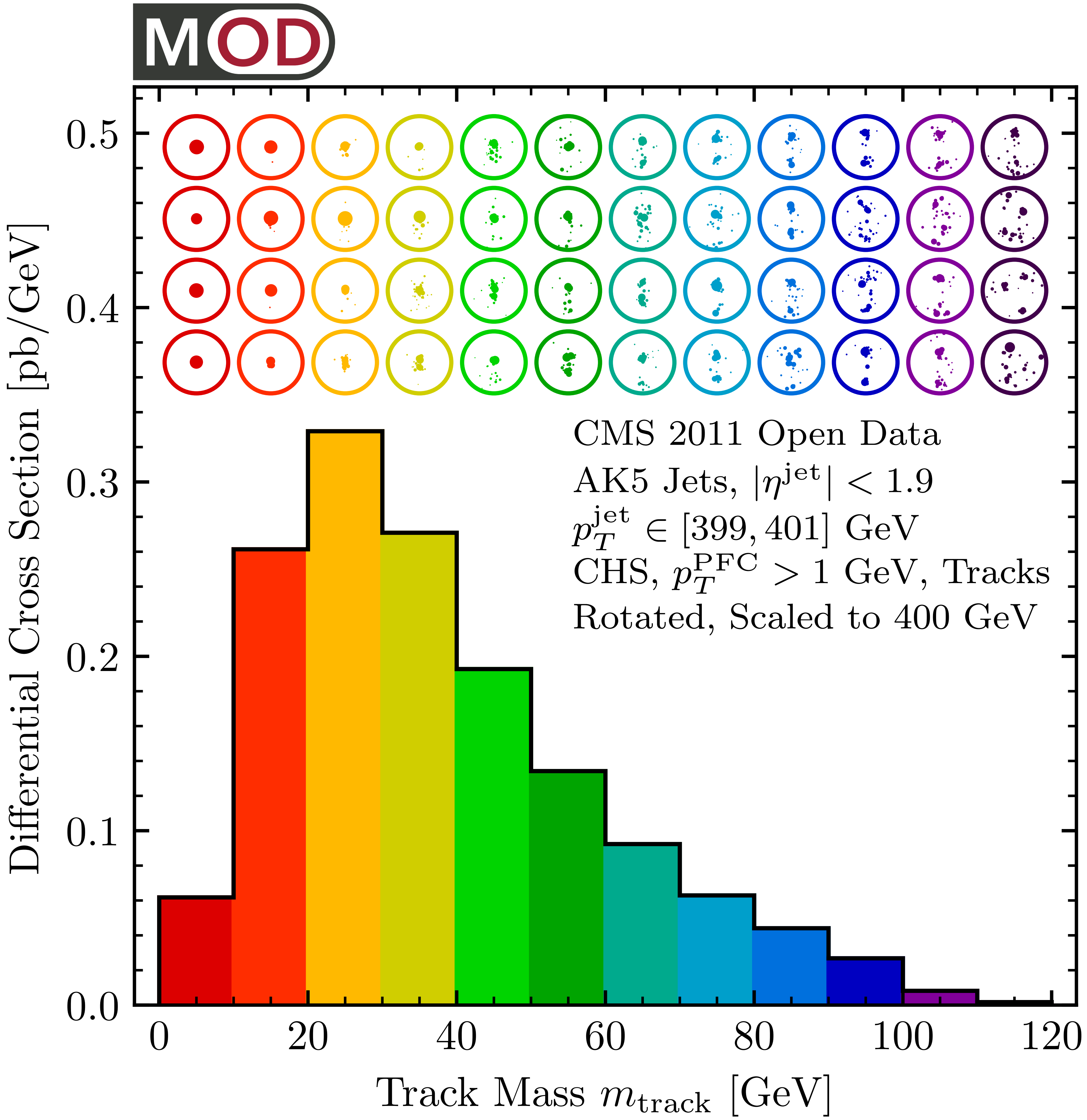
QCD produces mostly one-pronged jets but has long tails

Visualizing Geometry in CMS Open Data

[PTK, Mastandrea, Metodiev, Naik, Thaler, PRD 2019; code and datasets at energyflow.network]



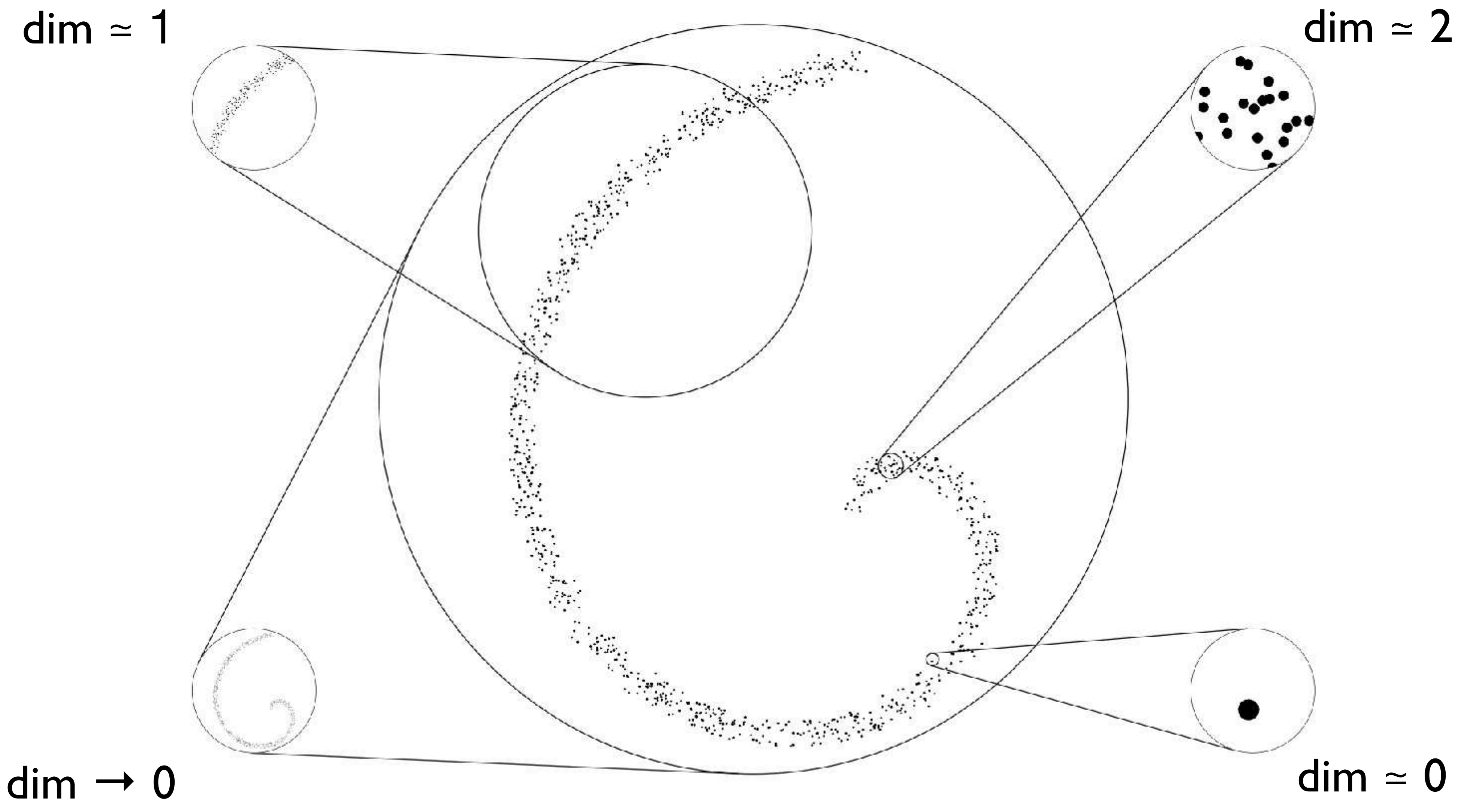
QCD produces mostly one-pronged jets but has long tails



4 most representative jets (medoids) shown for each bin

Quantifying Event-Space Manifolds

Correlation dimension: *how does the # of elements within a ball of size Q change?*



$$N_{\text{neigh.}}(Q) \propto Q^{\text{dim}} \implies \text{dim}(Q) = Q\frac{d}{dQ}\ln N_{\text{neigh.}}(Q)$$

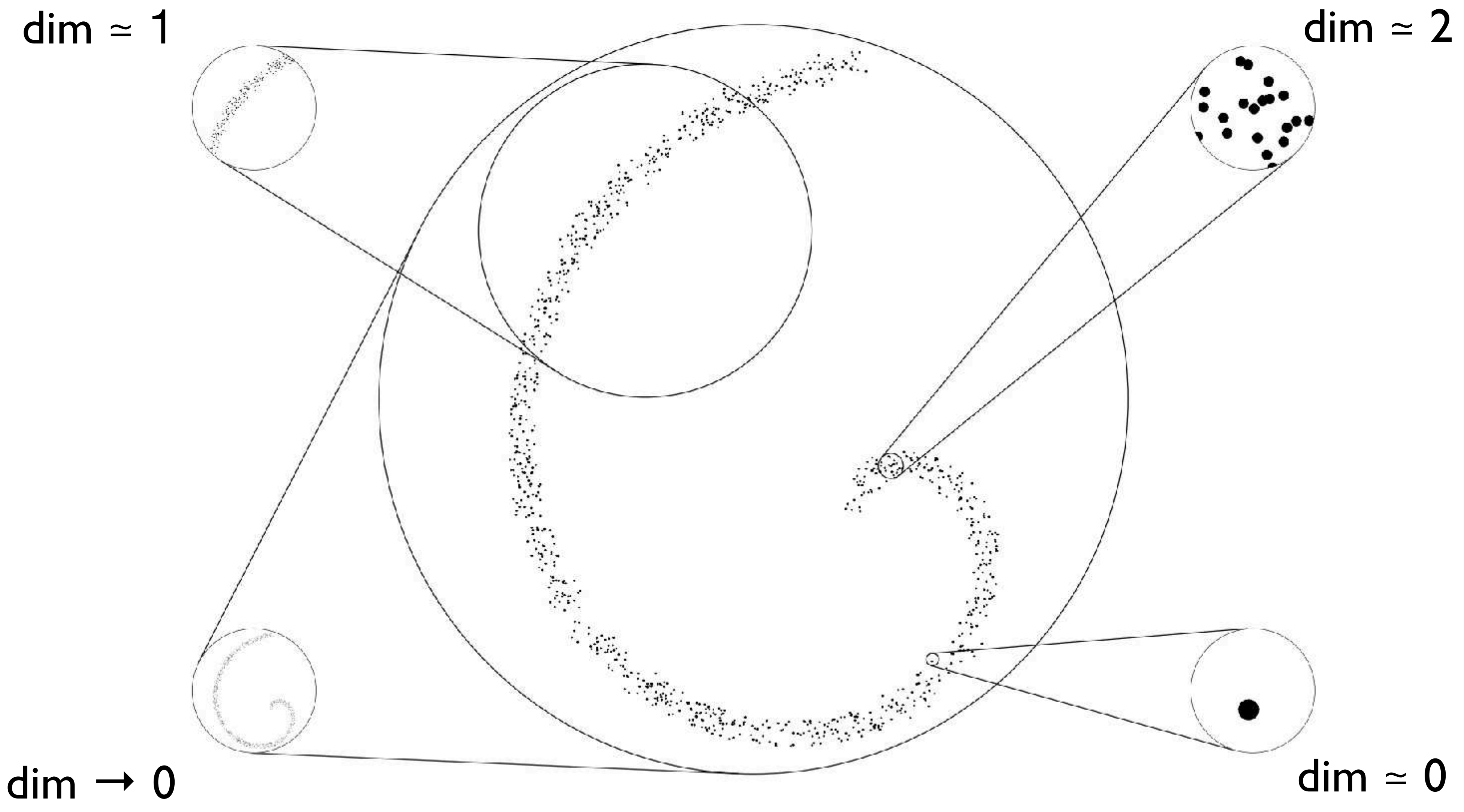
[Grassberger, Procaccia, PRL 1983; PTK, Metodiev, Thaler, PRL 2019]

*Preliminary LL calculation for QG jets in backup

Quantifying Event-Space Manifolds

$$\dim(Q) = Q\frac{\partial}{\partial Q}\ln\sum_i\sum_j\Theta(\mathrm{EMD}(\mathcal{E}_i,\mathcal{E}'_j) < Q)$$

Correlation dimension: *how does the # of elements within a ball of size Q change?*



$$N_{\text{neigh.}}(Q) \propto Q^{\dim} \implies \dim(Q) = Q\frac{d}{dQ}\ln N_{\text{neigh.}}(Q)$$

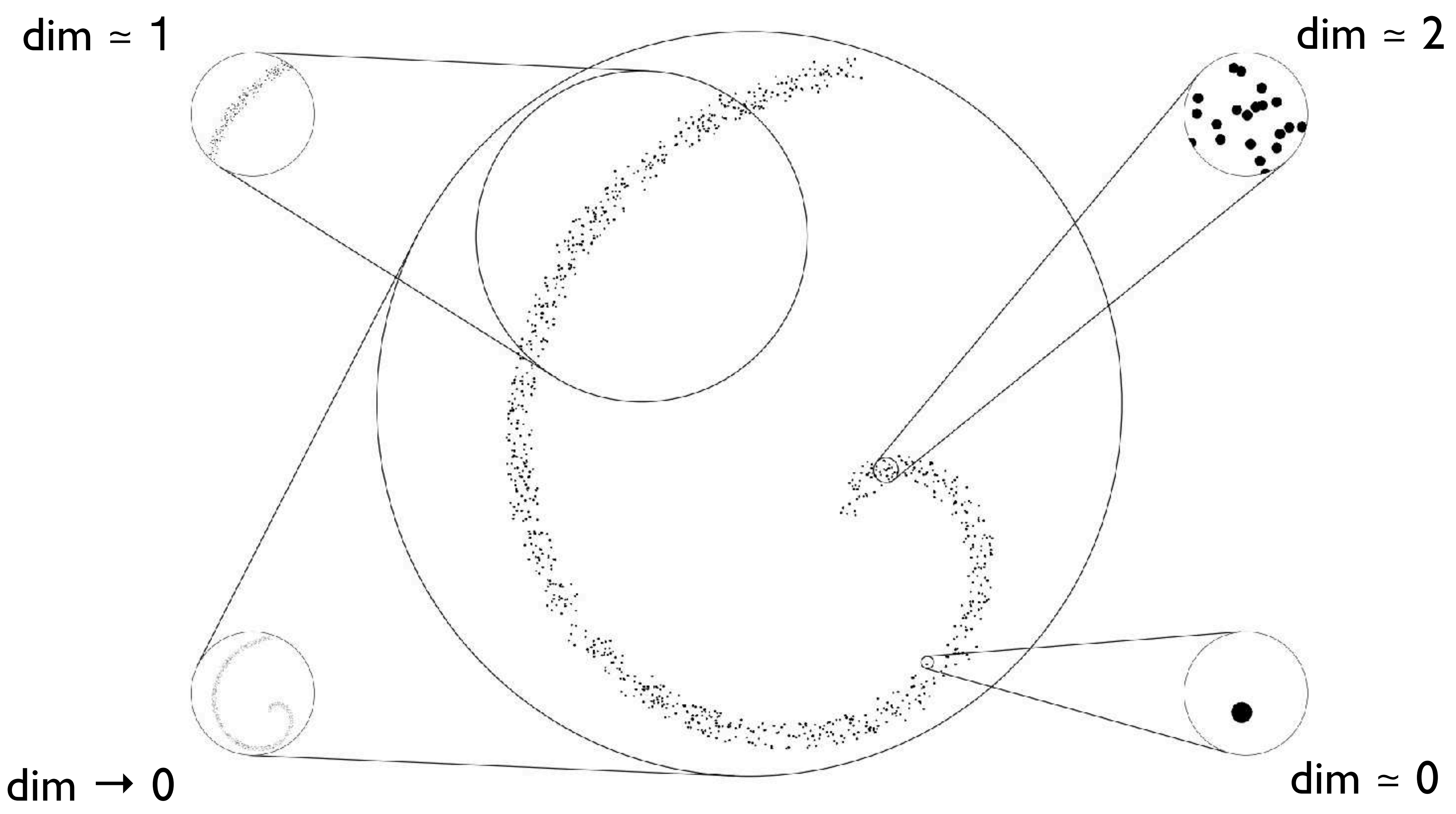
[Grassberger, Procaccia, PRL 1983; PTK, Metodiev, Thaler, PRL 2019]

*Preliminary LL calculation for QG jets in backup

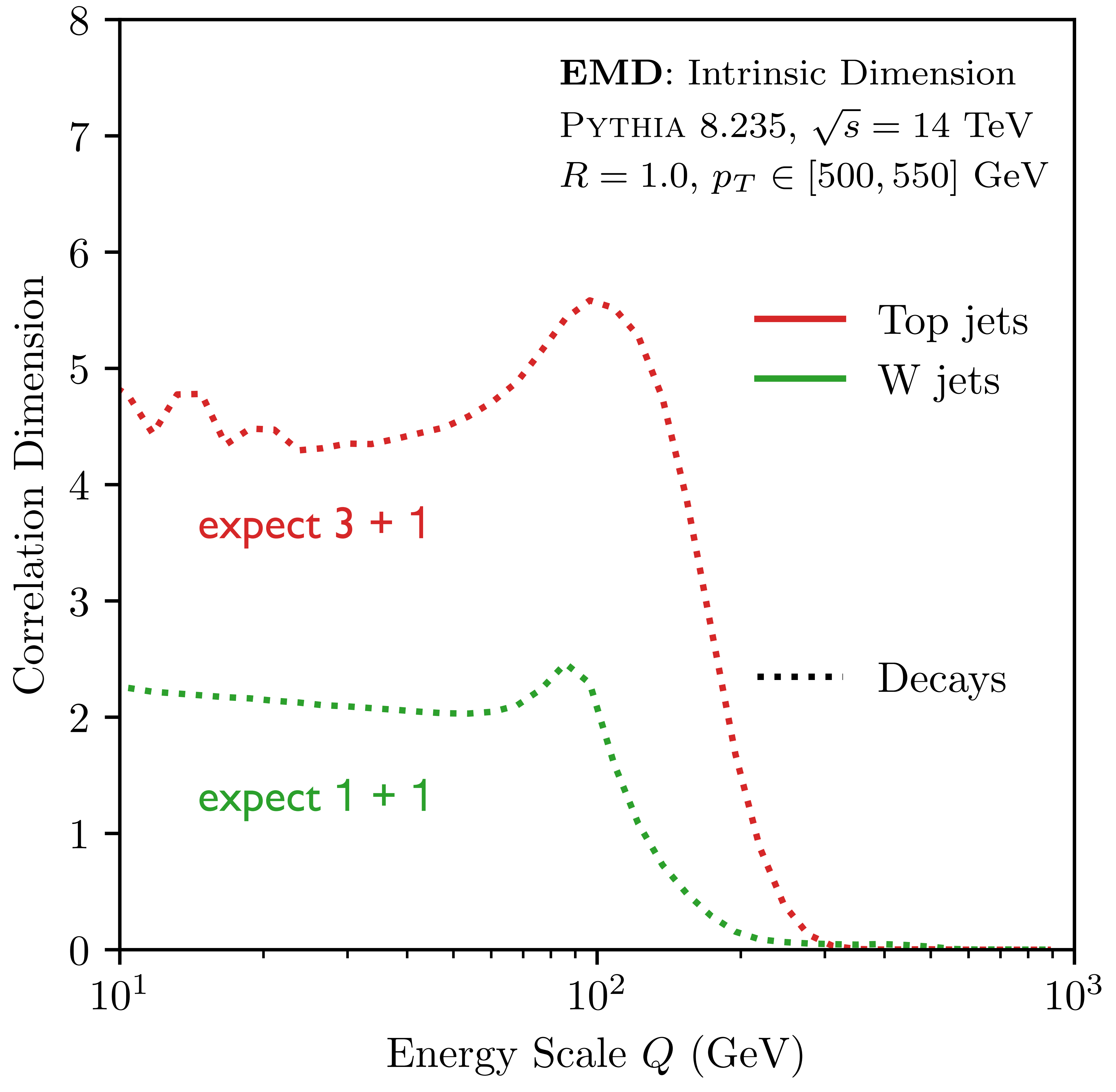
Quantifying Event-Space Manifolds

$$\dim(Q) = Q \frac{\partial}{\partial Q} \ln \sum_i \sum_j \Theta(\text{EMD}(\mathcal{E}_i, \mathcal{E}'_j) < Q)$$

Correlation dimension: *how does the # of elements within a ball of size Q change?*



$$N_{\text{neigh.}}(Q) \propto Q^{\text{dim}} \implies \dim(Q) = Q \frac{d}{dQ} \ln N_{\text{neigh.}}(Q)$$



Correlation dimension lessons:

Decays are "constant" dim. at low *Q*

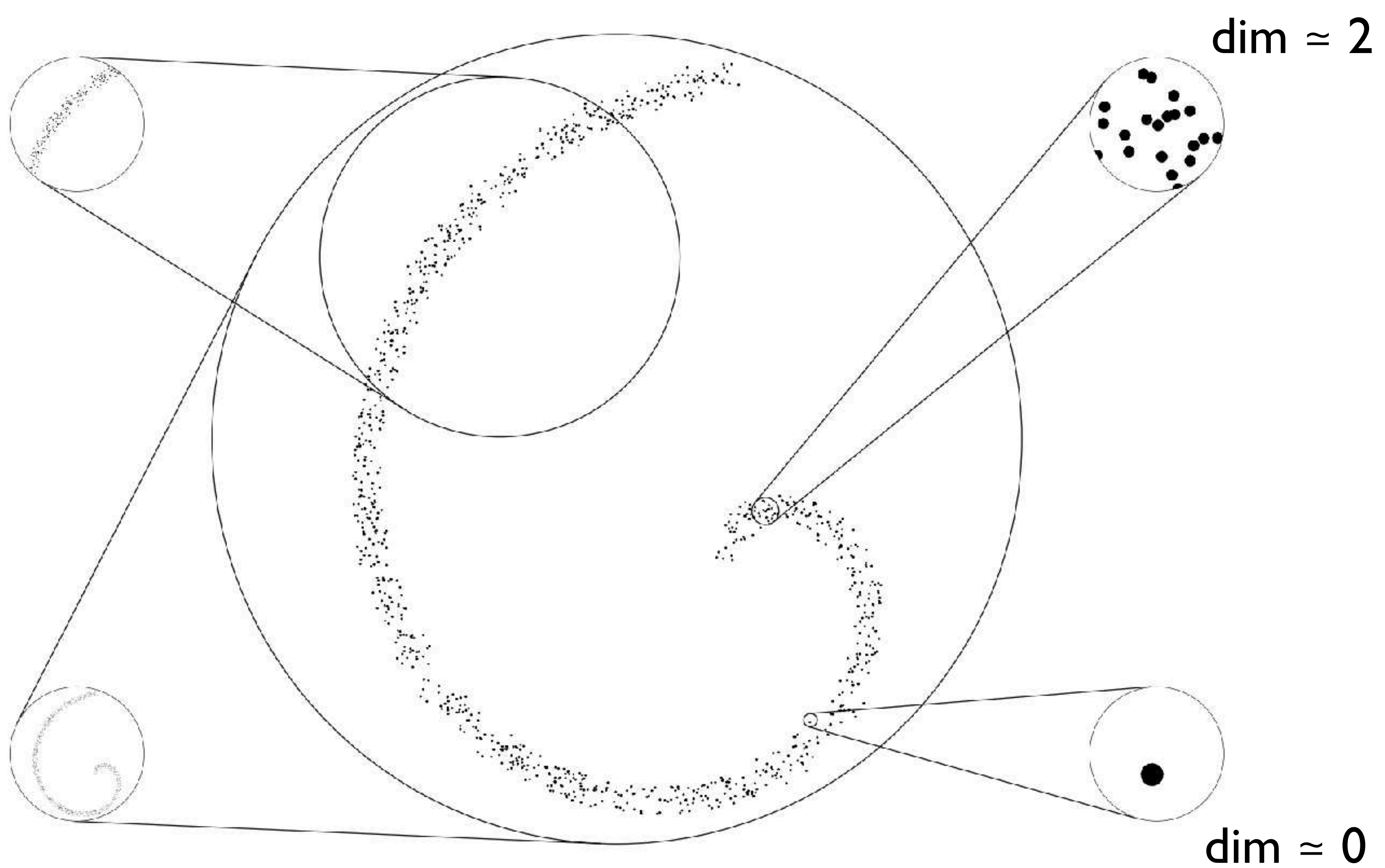
[Grassberger, Procaccia, PRL 1983; PTK, Metodiev, Thaler, PRL 2019]

*Preliminary LL calculation for QG jets in backup

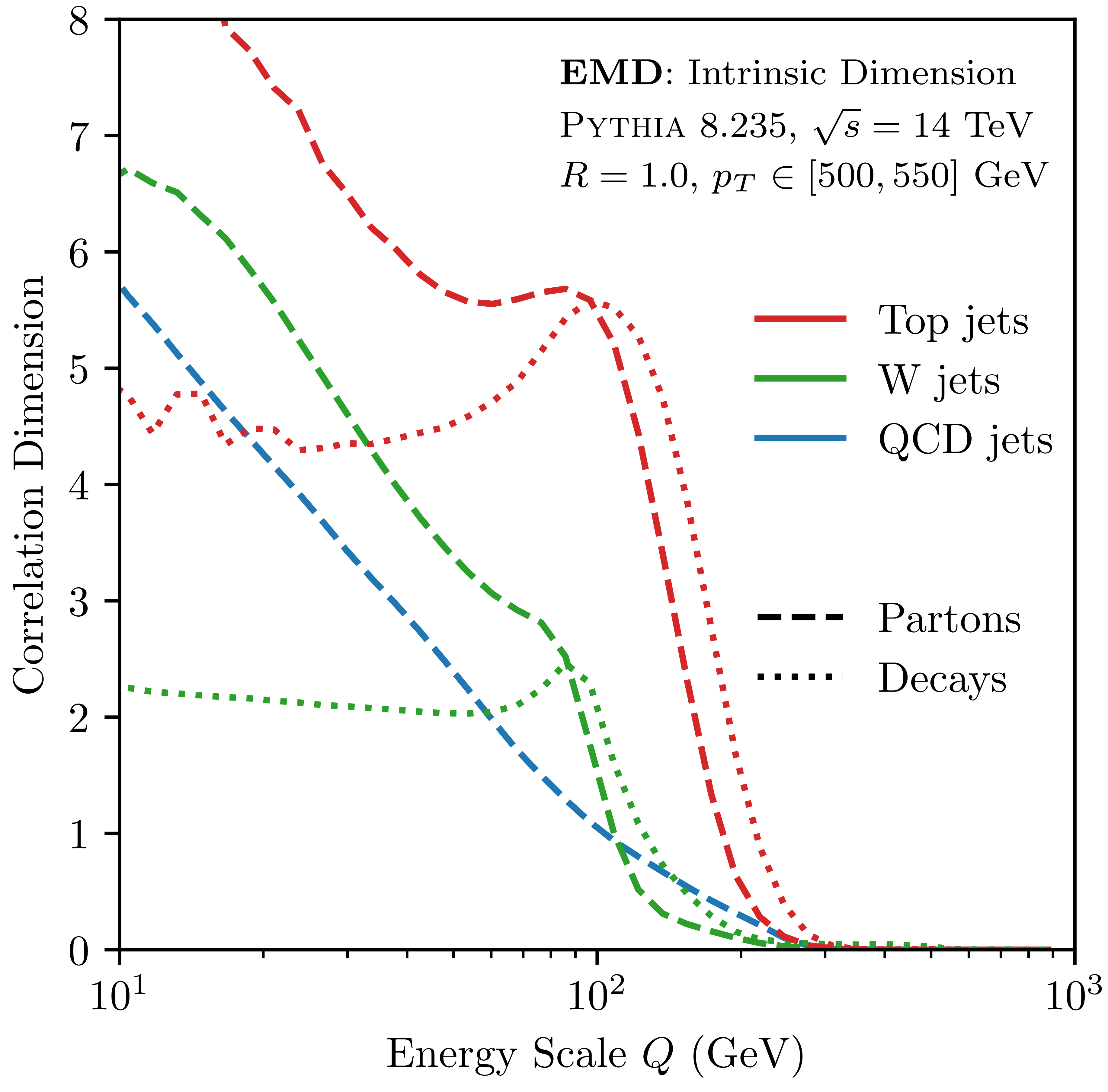
Quantifying Event-Space Manifolds

$$\dim(Q) = Q\frac{\partial}{\partial Q}\ln\sum_i\sum_j\Theta(\mathrm{EMD}(\mathcal{E}_i,\mathcal{E}'_j) < Q)$$

Correlation dimension: *how does the # of elements within a ball of size Q change?*



$$N_{\text{neigh.}}(Q) \propto Q^{\dim} \implies \dim(Q) = Q\frac{d}{dQ}\ln N_{\text{neigh.}}(Q)$$



Correlation dimension lessons:

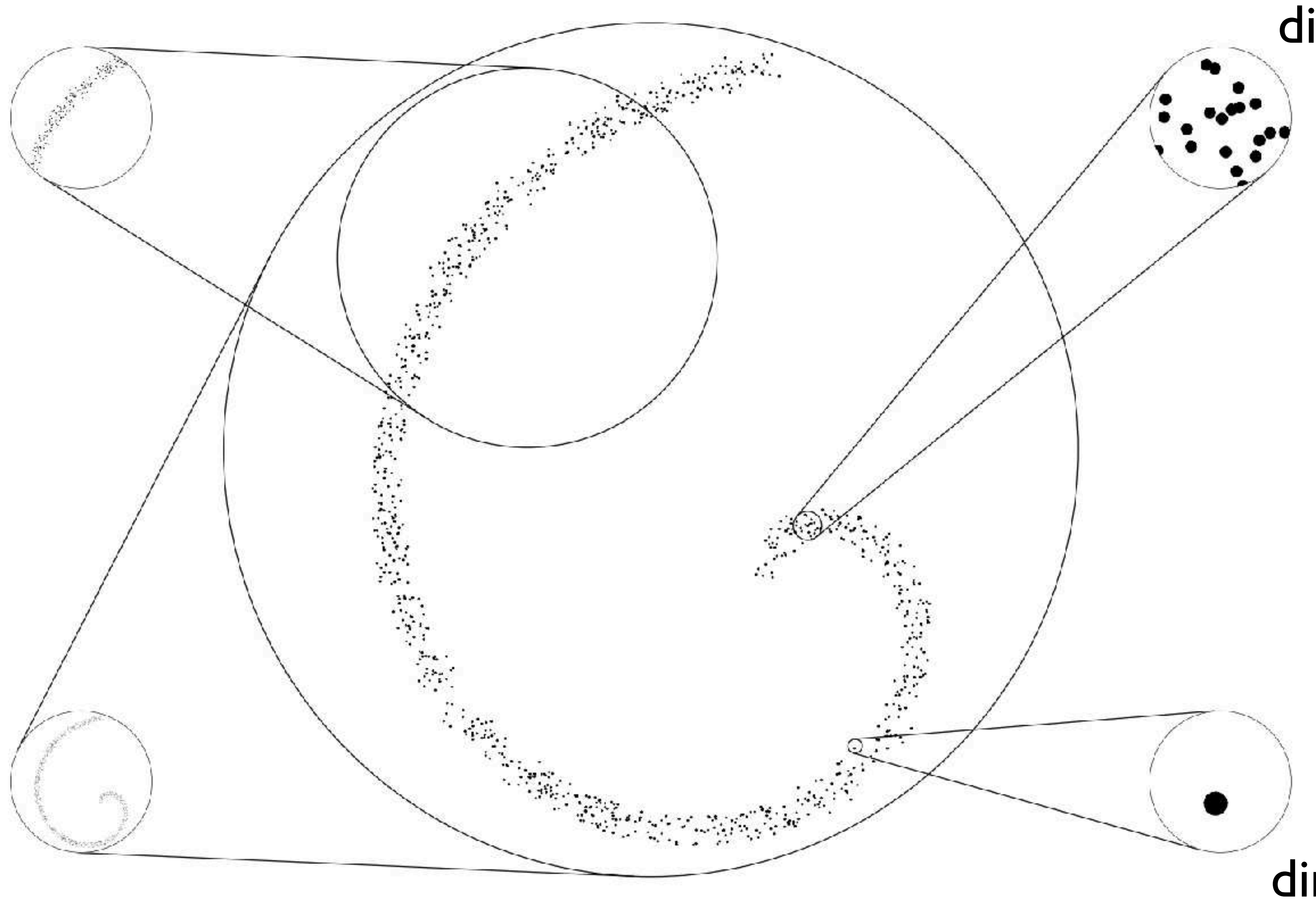
- Decays are "constant" dim. at low Q
- Complexity hierarchy: QCD < W < Top
- Fragmentation increases dim. at smaller scales

[Grassberger, Procaccia, PRL 1983; PTK, Metodiev, Thaler, PRL 2019]

*Preliminary LL calculation for QG jets in backup

Quantifying Event-Space Manifolds

Correlation dimension: *how does the # of elements within a ball of size Q change?*

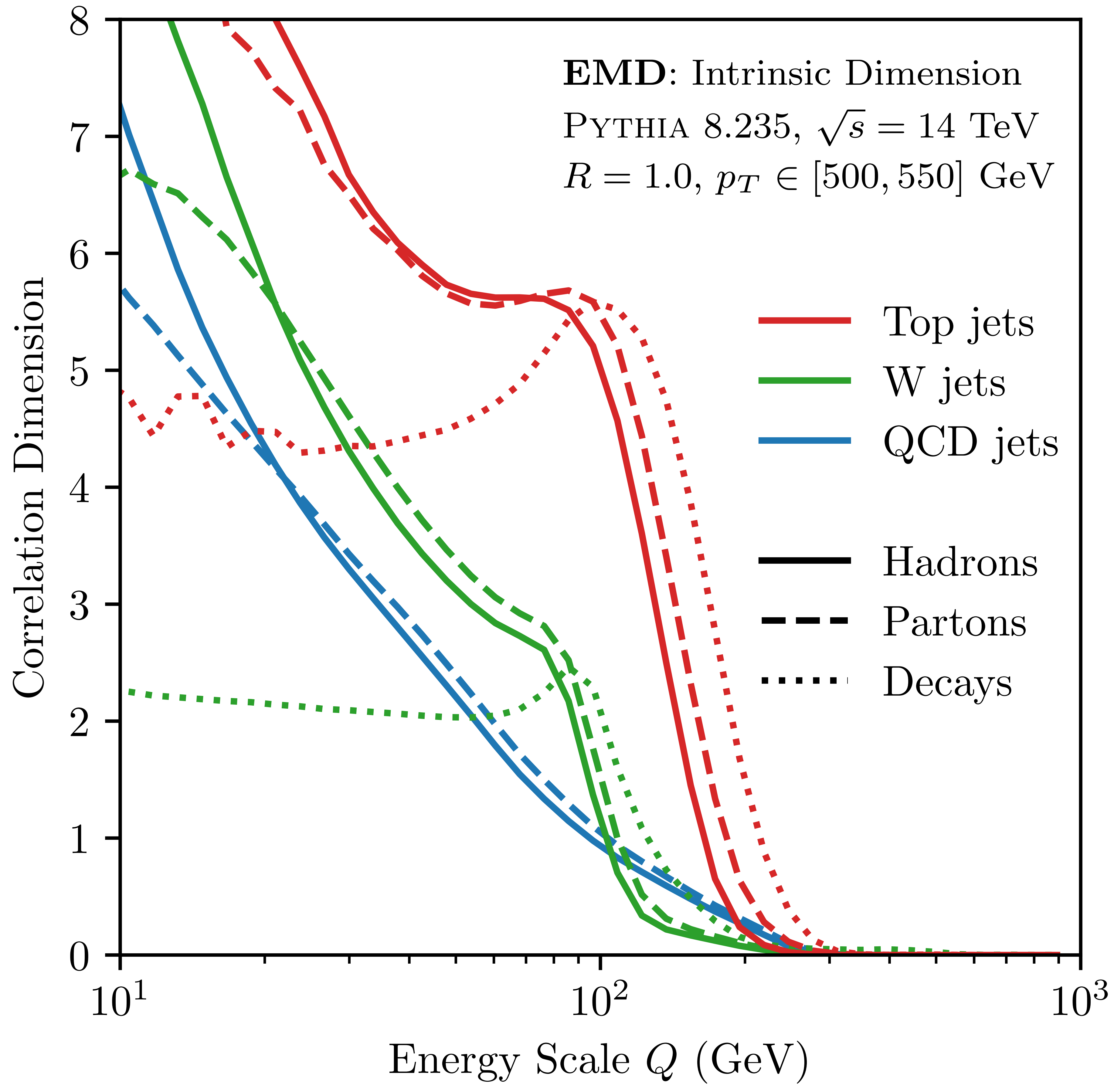


$$N_{\text{neigh.}}(Q) \propto Q^{\text{dim}} \implies \dim(Q) = Q\frac{d}{dQ}\ln N_{\text{neigh.}}(Q)$$

Correlation dimension lessons:

- Decays are "constant" dim. at low Q
- Complexity hierarchy: QCD < W < Top
- Fragmentation increases dim. at smaller scales
- Hadronization important around 20-30 GeV

$$\dim(Q) = Q\frac{\partial}{\partial Q}\ln \sum_i \sum_j \Theta(\text{EMD}(\mathcal{E}_i, \mathcal{E}'_j) < Q)$$

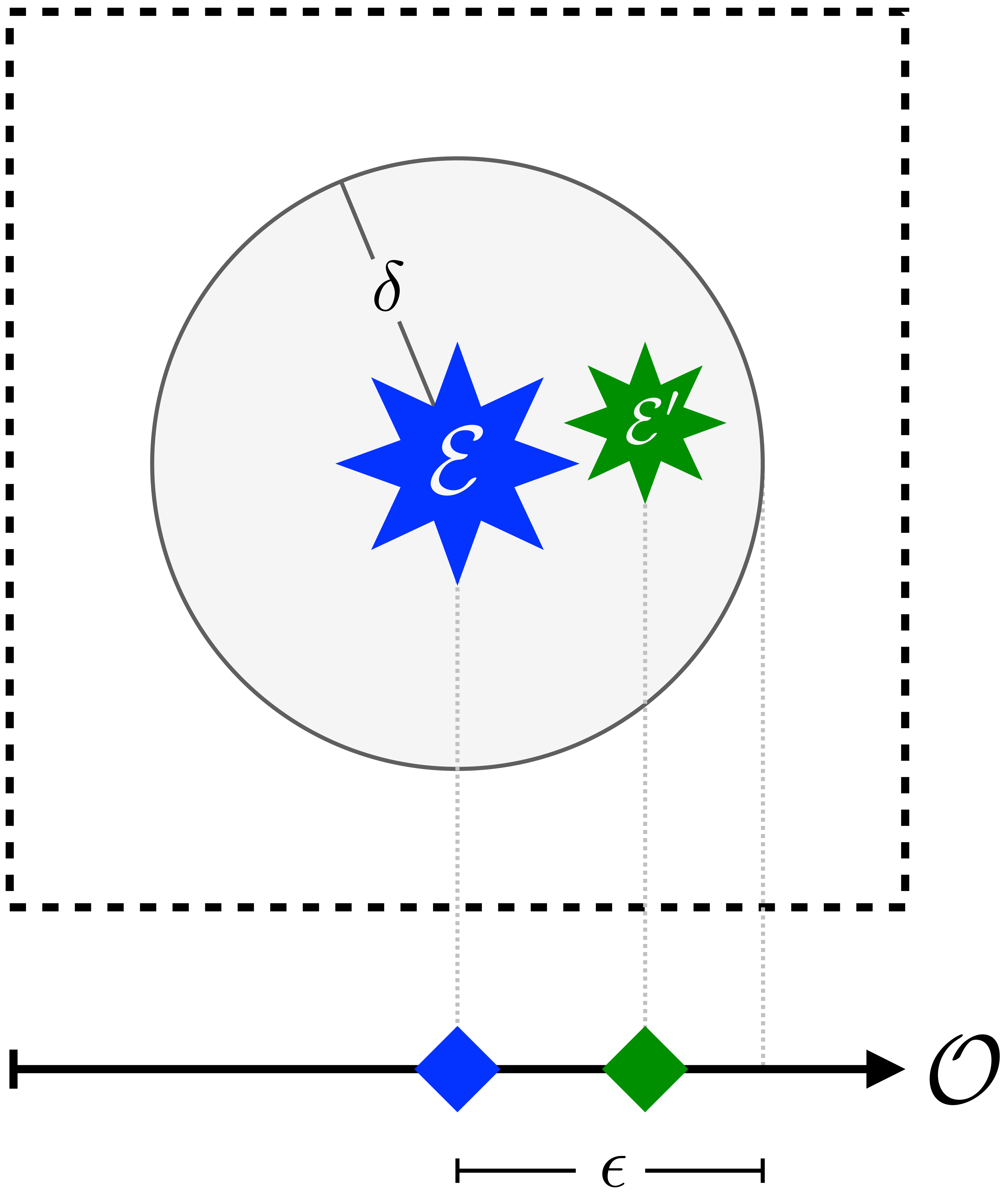


[Grassberger, Procaccia, PRL 1983; PTK, Metodiev, Thaler, PRL 2019]

*Preliminary LL calculation for QG jets in backup

More EMD Geometry – Continuity in the Space of Events

[PTK, Metodiev, Thaler, 2004.04159]



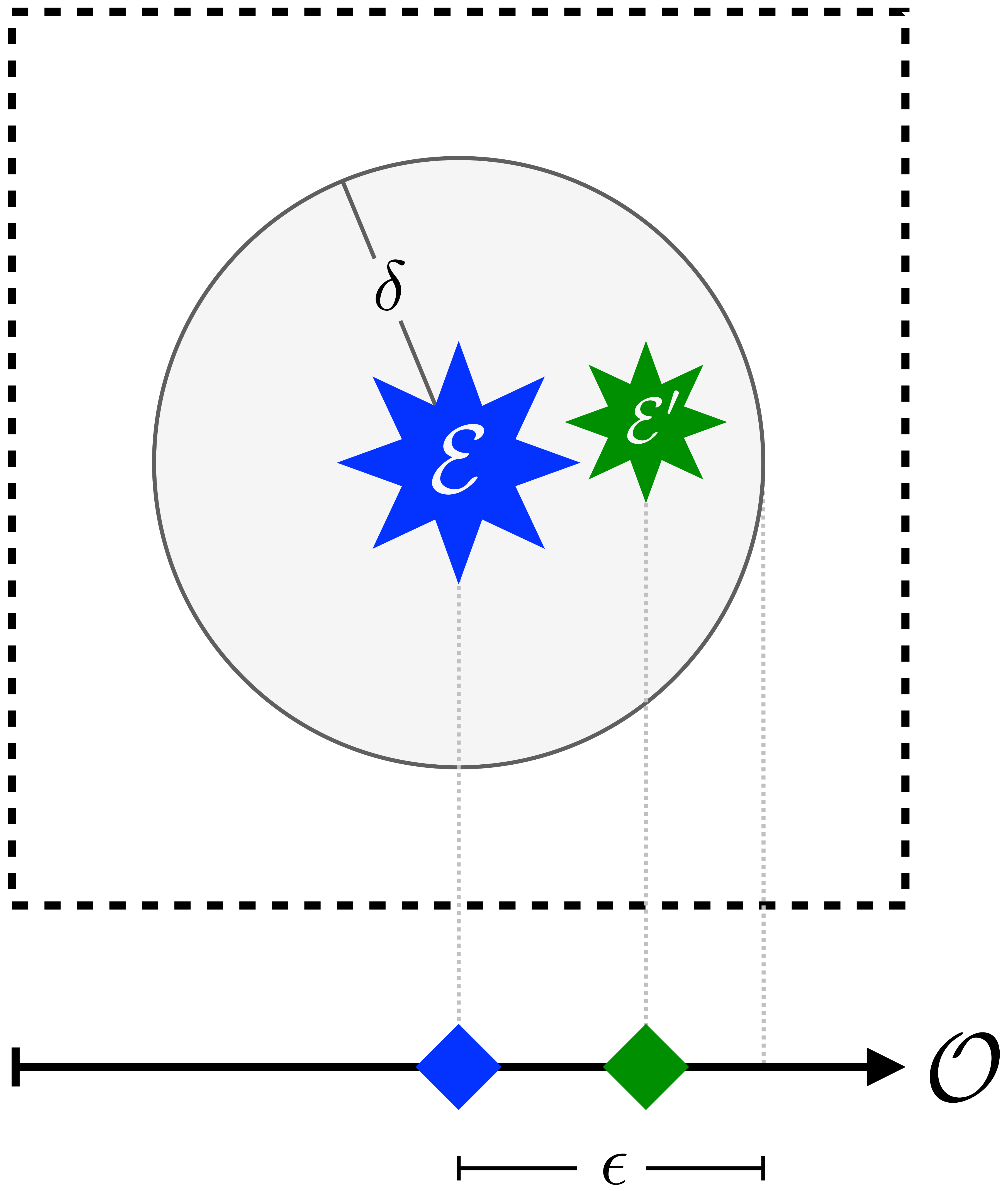
Classic ϵ – δ definition of continuity in a metric space

An observable $\mathcal{O}$ is EMD continuous at an event $\mathcal{E}$ if, for any $\epsilon > 0$, there exists a $\delta > 0$ such that for all events $\mathcal{E}'$:

$$\text{EMD}(\mathcal{E}, \mathcal{E}') < \delta \quad \Longrightarrow \quad |\mathcal{O}(\mathcal{E}) - \mathcal{O}(\mathcal{E}')| < \epsilon.$$

More EMD Geometry – Continuity in the Space of Events

[PTK, Metodiev, Thaler, 2004.04159]



Classic $\epsilon - \delta$ definition of continuity in a metric space

An observable $\mathcal{O}$ is EMD continuous at an event $\mathcal{E}$ if, for any $\epsilon > 0$, there exists a $\delta > 0$ such that for all events $\mathcal{E}'$:

$$\text{EMD}(\mathcal{E}, \mathcal{E}') < \delta \implies |\mathcal{O}(\mathcal{E}) - \mathcal{O}(\mathcal{E}')| < \epsilon.$$

Towards a geometric definition of IRC Safety

IRC Safety = EMD Continuity*

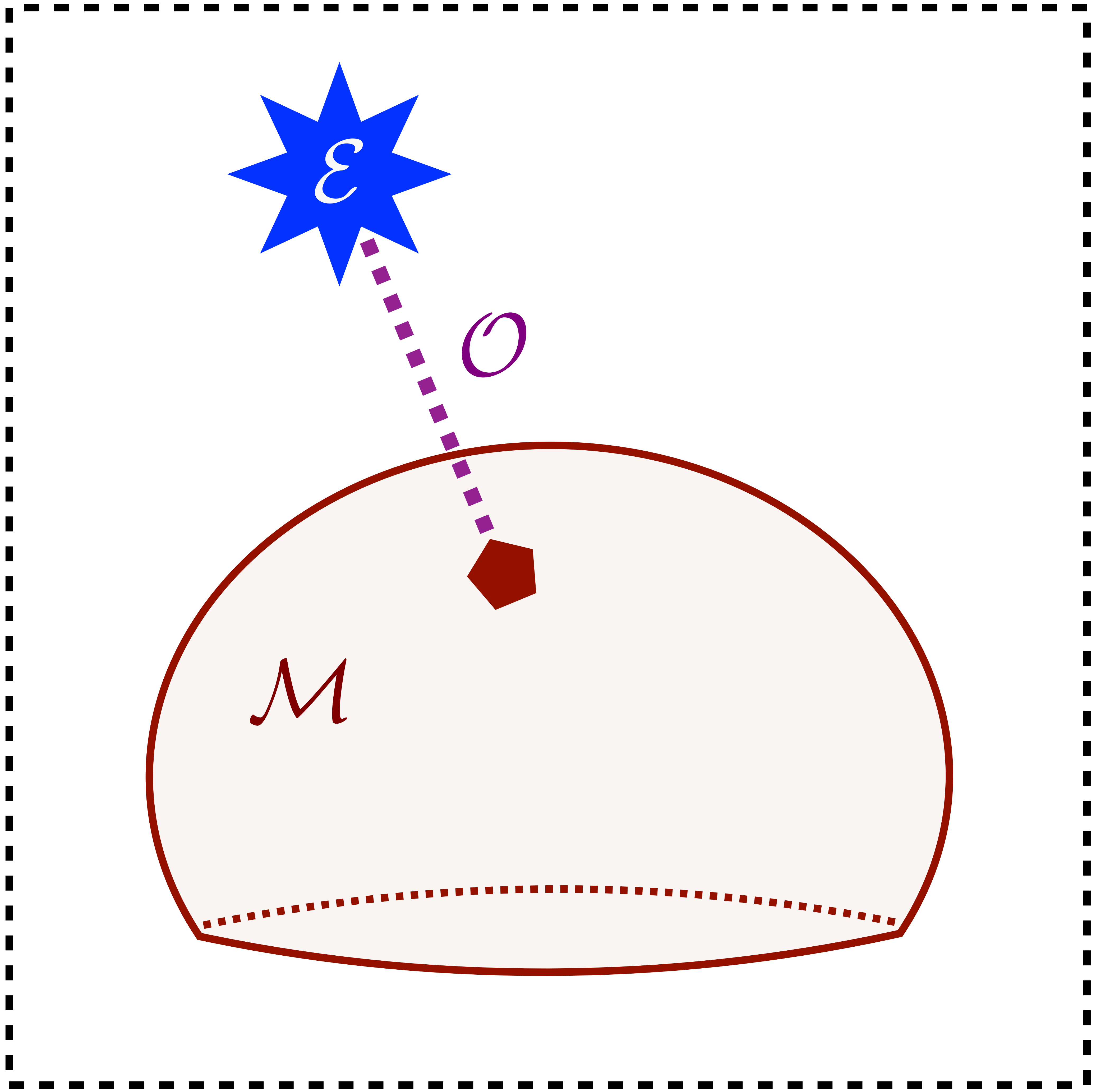
*on all but a negligible set‡ of events

‡a negligible set is one that contains no positive-radius EMD-ball

⋮

Defining Observables via Event Space Geometry

[PTK, Metodiev, Thaler, 2004.04159]

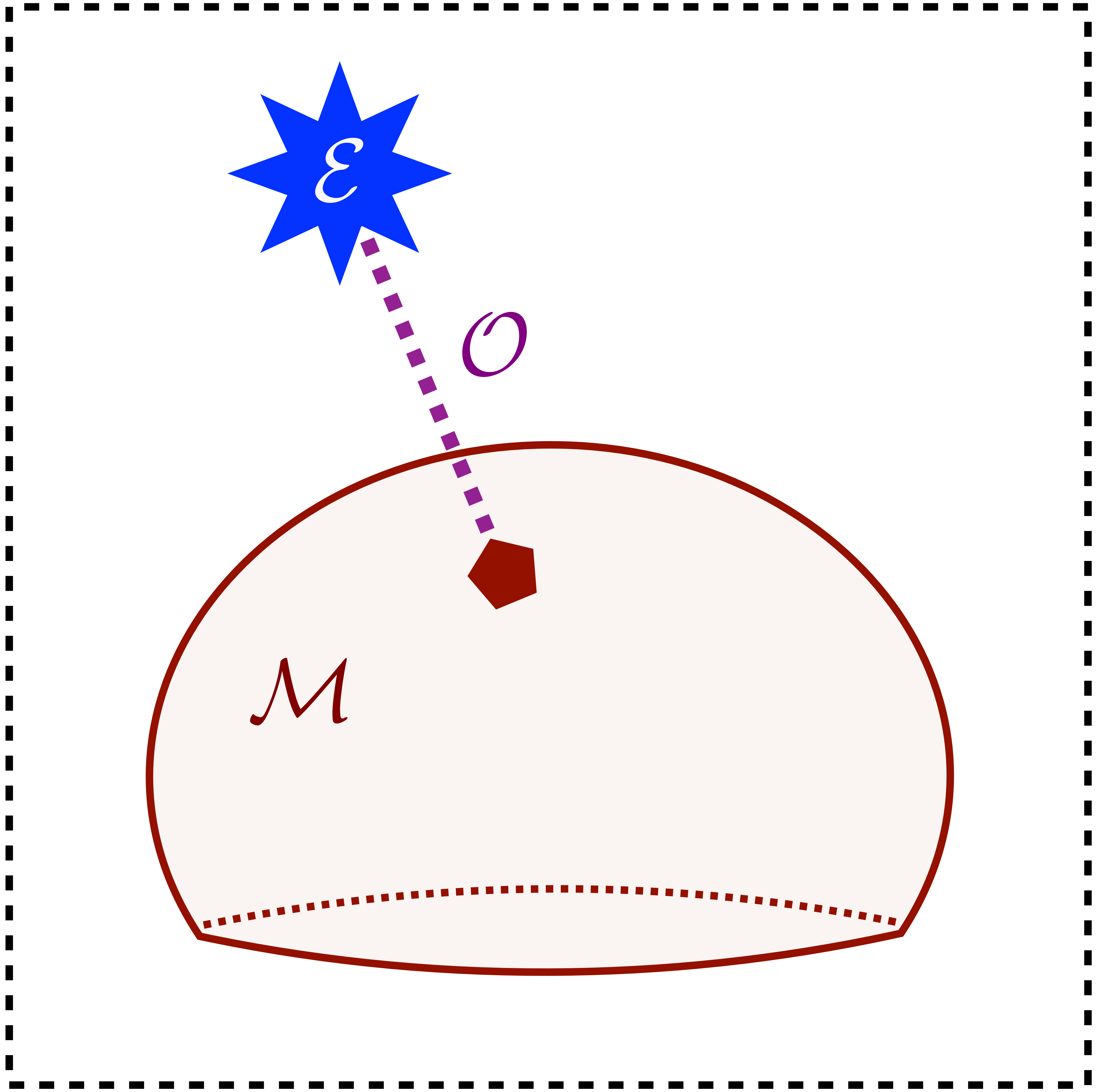


Many common observables are distance of closest approach from event to a specific manifold

$$\mathcal{O}(\mathcal{E}) = \min_{\mathcal{E}' \in \mathcal{M}} \mathrm{EMD}_{\beta,R}(\mathcal{E}, \mathcal{E}')$$

Defining Observables via Event Space Geometry

[PTK, Metodiev, Thaler, 2004.04159]



Many common observables are distance of closest approach from event to a specific manifold

$$\mathcal{O}(\mathcal{E}) = \min_{\mathcal{E}' \in \mathcal{M}} \text{EMD}_{\beta,R}(\mathcal{E}, \mathcal{E}')$$

EMD variant for equal-energy events

$$\text{EMD}_\beta(\mathcal{E}, \mathcal{E}') = \lim_{R\to\infty} R^\beta \text{EMD}_{\beta,R}(\mathcal{E}, \mathcal{E}') = \min_{\{f_{ij}\geq 0\}} \sum_{i=1}^{M} \sum_{j=1}^{M'} f_{ij}\theta_{ij}^\beta$$

Enforces equal energy (else infinity)

on equal-energy events

Defining Observables via Event Space Geometry

$$\mathcal{O}(\mathcal{E}) = \min_{\mathcal{E}' \in \mathcal{M}} \text{EMD}_{\beta,R}(\mathcal{E}, \mathcal{E}')$$

[PTK, Metodiev, Thaler, 2004.04159]

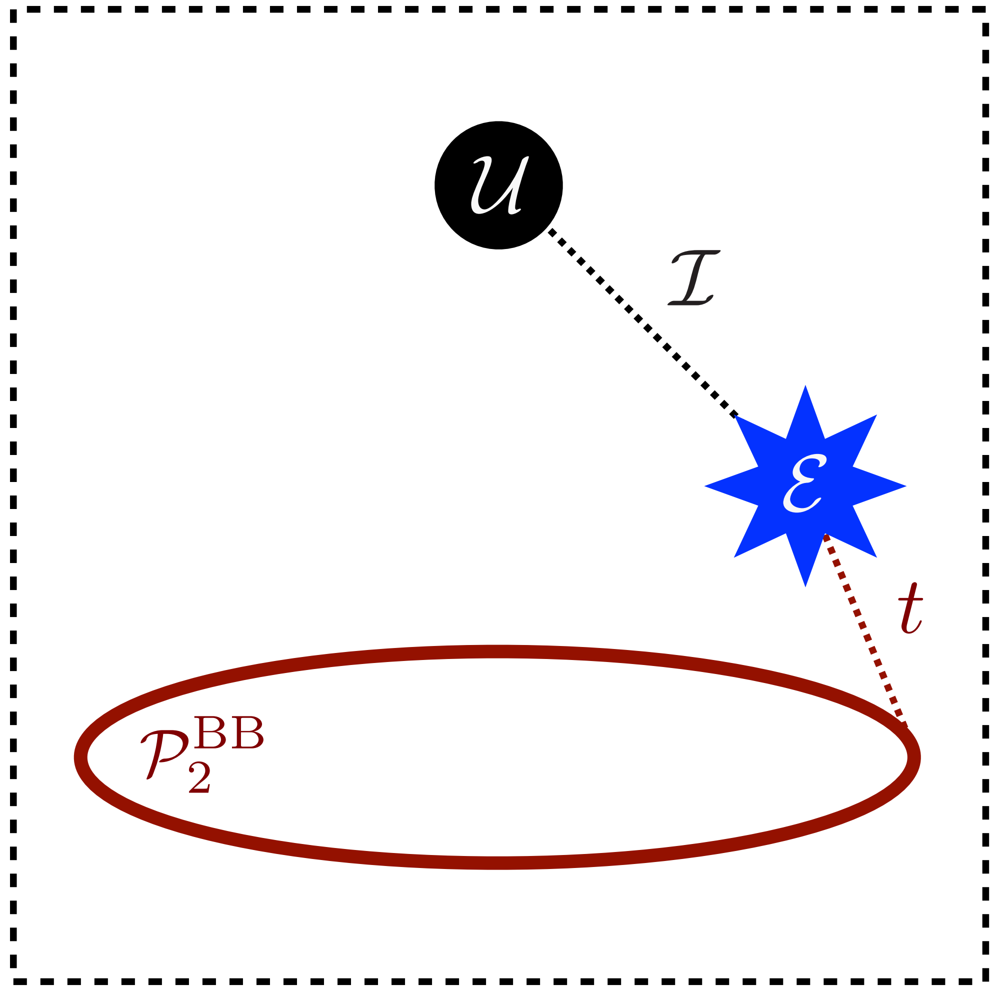
Defining Observables via Event Space Geometry

$$\mathcal{O}(\mathcal{E}) = \min_{\mathcal{E}' \in \mathcal{M}} \mathrm{EMD}_{\beta,R}(\mathcal{E}, \mathcal{E}')$$

[PTK, Metodiev, Thaler, 2004.04159]

Thrust, spherocity, isotropy*

Distance of closest approach to a specific manifold



$$t(\mathcal{E}) = \min_{\mathcal{E}' \in \mathcal{P}_2^{\mathrm{BB}}} \mathrm{EMD}_2(\mathcal{E}, \mathcal{E}')$$

$$\sqrt{s(\mathcal{E})} = \min_{\mathcal{E}' \in \mathcal{P}_2^{\mathrm{BB}}} \mathrm{EMD}_1(\mathcal{E}, \mathcal{E}')$$

$$\mathcal{I}^{(\beta)}(\mathcal{E}) = \min_{\mathcal{E}' \in M_{\mathcal{U}}} \mathrm{EMD}_\beta(\mathcal{E}, \mathcal{E}')$$

[Farhi, PRL 1977; Georgi, Machacek, PRL 1977]
*New! [Cesarotti, Thaler, 2004.06125]

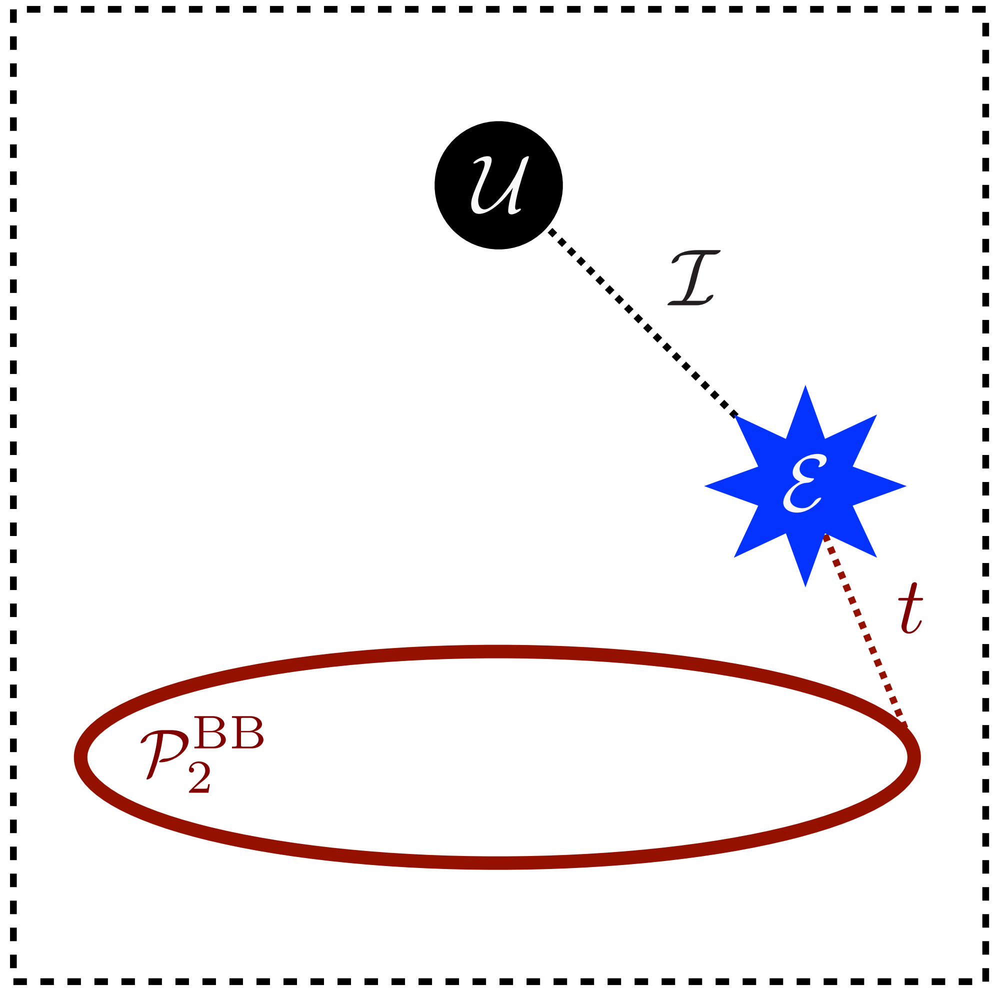
Defining Observables via Event Space Geometry

$$\mathcal{O}(\mathcal{E}) = \min_{\mathcal{E}' \in \mathcal{M}} \text{EMD}_{\beta,R}(\mathcal{E}, \mathcal{E}')$$

[PTK, Metodiev, Thaler, 2004.04159]

Thrust, spherocity, isotropy*

Distance of closest approach to a specific manifold



$$t(\mathcal{E}) = \min_{\mathcal{E}' \in \mathcal{P}_2^{\text{BB}}} \text{EMD}_2(\mathcal{E}, \mathcal{E}')$$

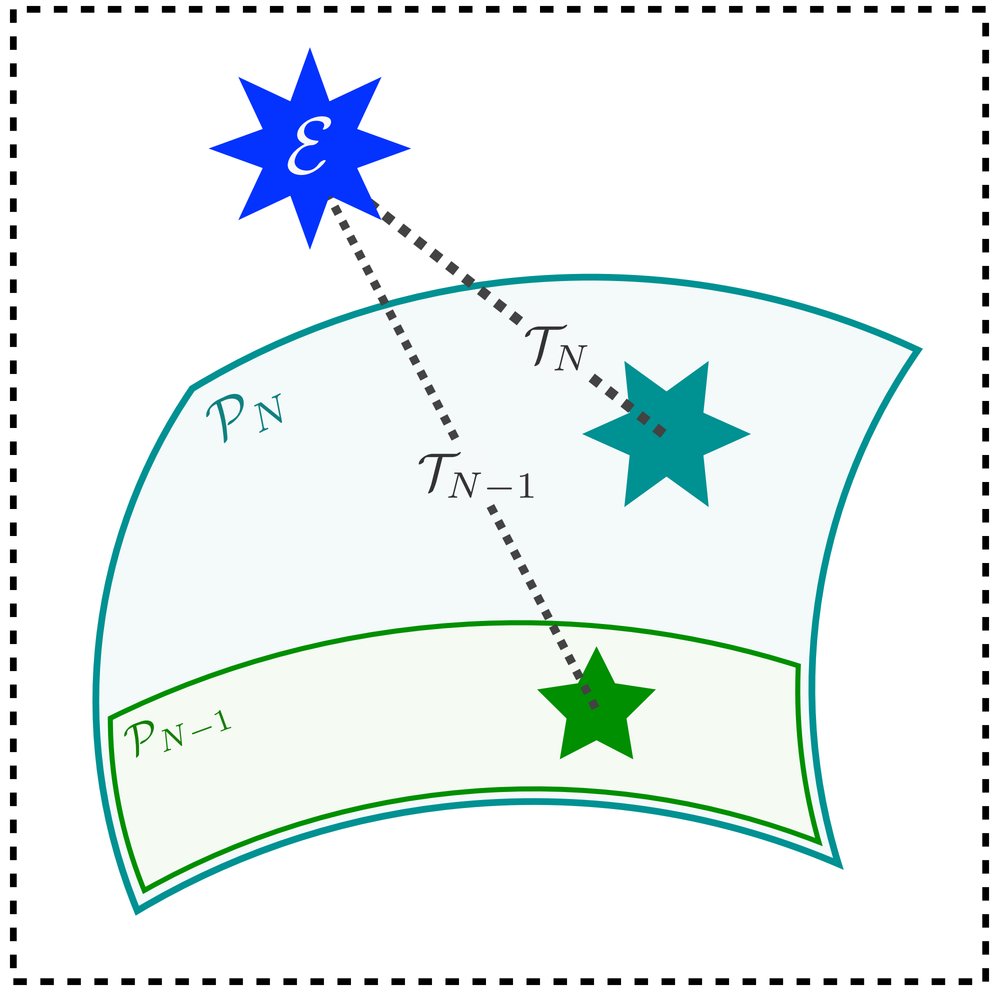
$$\sqrt{s(\mathcal{E})} = \min_{\mathcal{E}' \in \mathcal{P}_2^{\text{BB}}} \text{EMD}_1(\mathcal{E}, \mathcal{E}')$$

$$\mathcal{I}^{(\beta)}(\mathcal{E}) = \min_{\mathcal{E}' \in M_{\mathcal{U}}} \text{EMD}_\beta(\mathcal{E}, \mathcal{E}')$$

[Farhi, PRL 1977; Georgi, Machacek, PRL 1977]
*New! [Cesarotti, Thaler, 2004.06125]

N-jettiness

Minimum distance from event to N-particle manifold



without beam region

$$\mathcal{T}_N^{(\beta)}(\mathcal{E}) = \min_{\mathcal{E}' \in \mathcal{P}_N} \text{EMD}_\beta(\mathcal{E}, \mathcal{E}')$$

with constant beam distance R^β

$$\mathcal{T}_N^{(\beta,R)}(\mathcal{E}) = \min_{\mathcal{E}' \in \mathcal{P}_N} \text{EMD}_{\beta,R}(\mathcal{E}, \mathcal{E}')$$

[Brandt, Dahmen, Z. Phys 1979;
Stewart, Tackmann, Waalewijn, PRL 2010]

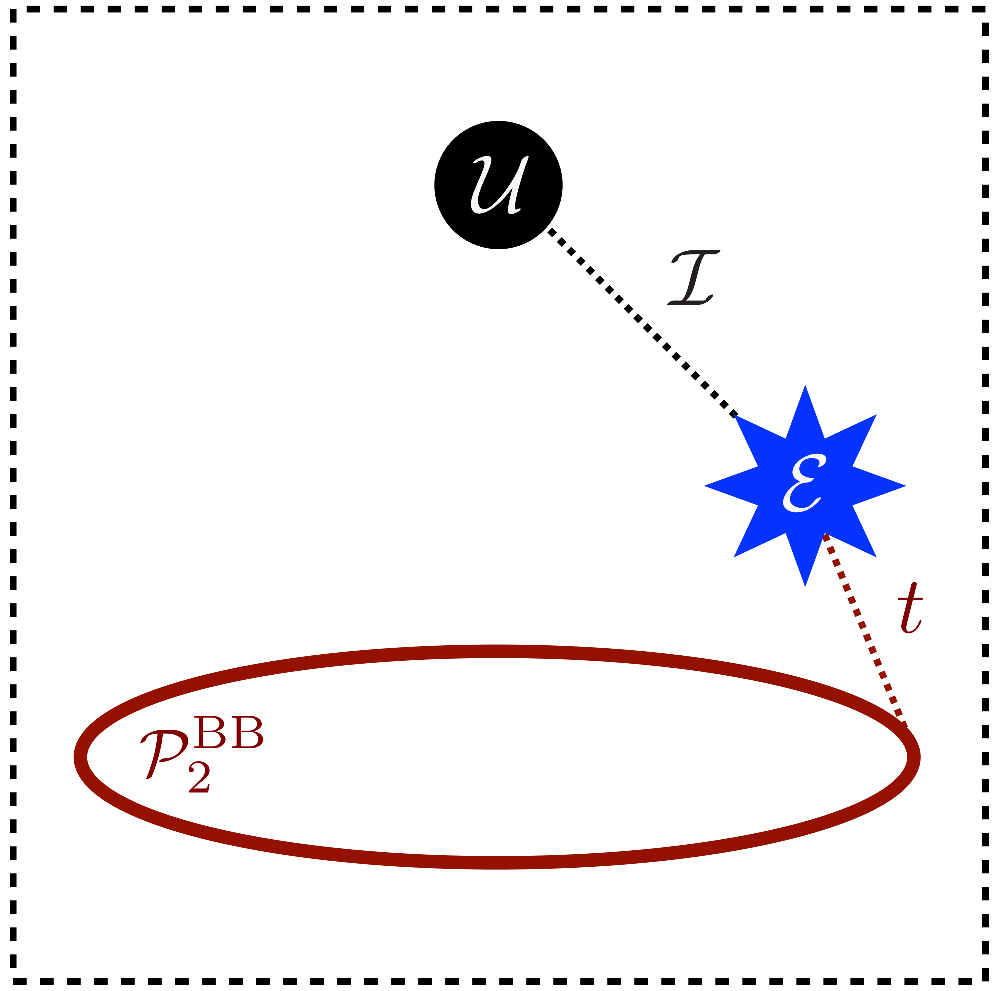
Defining Observables via Event Space Geometry

$$\mathcal{O}(\mathcal{E}) = \min_{\mathcal{E}' \in \mathcal{M}} \text{EMD}_{\beta,R}(\mathcal{E},\mathcal{E}')$$

[PTK, Metodiev, Thaler, 2004.04159]

Thrust, spherocity, isotropy*

Distance of closest approach to a specific manifold



$$t(\mathcal{E}) = \min_{\mathcal{E}' \in \mathcal{P}_2^{\text{BB}}} \text{EMD}_2(\mathcal{E},\mathcal{E}')$$

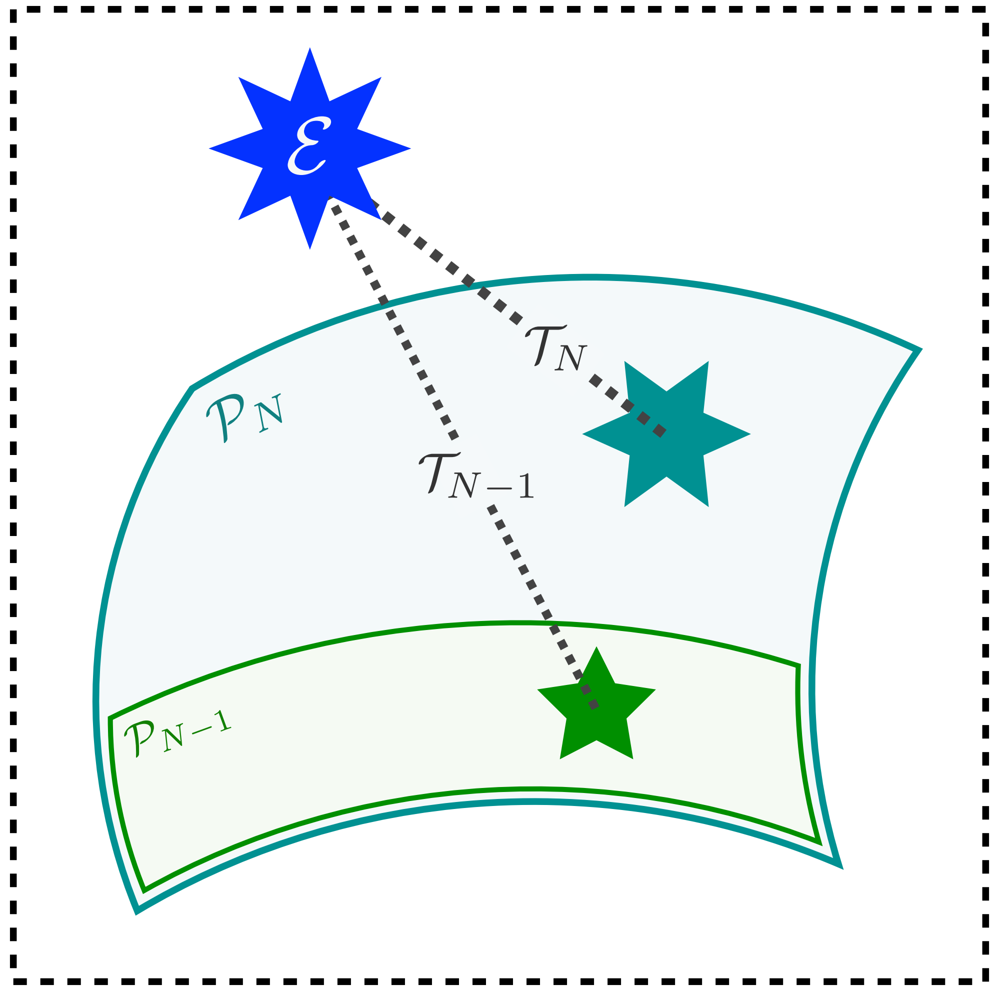
$$\sqrt{s(\mathcal{E})} = \min_{\mathcal{E}' \in \mathcal{P}_2^{\text{BB}}} \text{EMD}_1(\mathcal{E},\mathcal{E}')$$

$$\mathcal{I}^{(\beta)}(\mathcal{E}) = \min_{\mathcal{E}' \in M_{\mathcal{U}}} \text{EMD}_\beta(\mathcal{E},\mathcal{E}')$$

[Farhi, PRL 1977; Georgi, Machacek, PRL 1977]
*New! [Cesarotti, Thaler, 2004.06125]

N-jettiness

Minimum distance from event to N-particle manifold



without beam region

$$\mathcal{T}_N^{(\beta)}(\mathcal{E}) = \min_{\mathcal{E}' \in \mathcal{P}_N} \text{EMD}_\beta(\mathcal{E},\mathcal{E}')$$

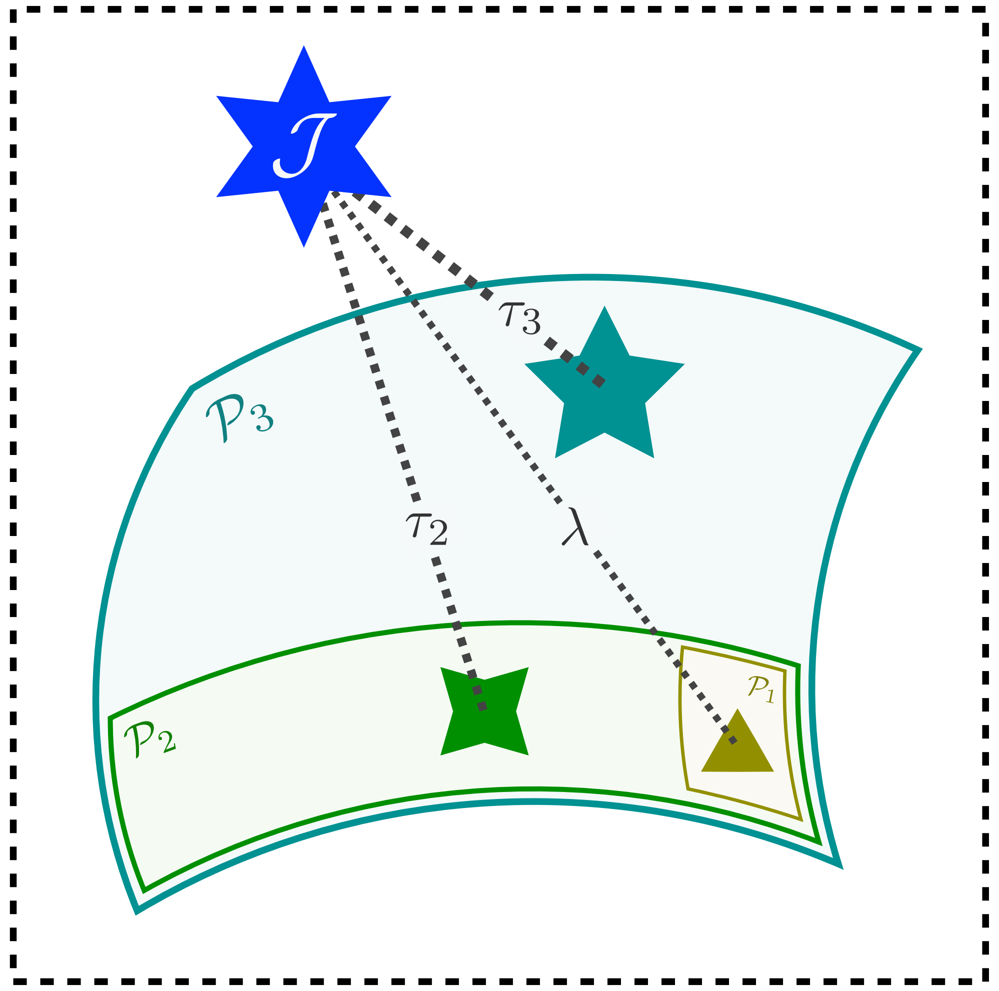
with constant beam distance R^β

$$\mathcal{T}_N^{(\beta,R)}(\mathcal{E}) = \min_{\mathcal{E}' \in \mathcal{P}_N} \text{EMD}_{\beta,R}(\mathcal{E},\mathcal{E}')$$

[Brandt, Dahmen, Z. Phys 1979;
Stewart, Tackmann, Waalewijn, PRL 2010]

N-subjettiness, angularities

Smallest distance from jet to N-particle manifold



for recoil-free angularity

$$\lambda_\beta(\mathcal{J}) = \min_{\mathcal{J}' \in \mathcal{P}_1} \text{EMD}_\beta(\mathcal{J},\mathcal{J}')$$

$$\tau_N^{(\beta)}(\mathcal{J}) = \min_{\mathcal{J}' \in \mathcal{P}_N} \text{EMD}_\beta(\mathcal{J},\mathcal{J}')$$

[Ellis, Vermilion, Walsh, Hornig, Lee, JHEP 2010;
Thaler, Van Tilburg, JHEP 2011, JHEP 2012]

Jets in the Space of Events – The Closest *N*-particle Description of an *M*-particle Event

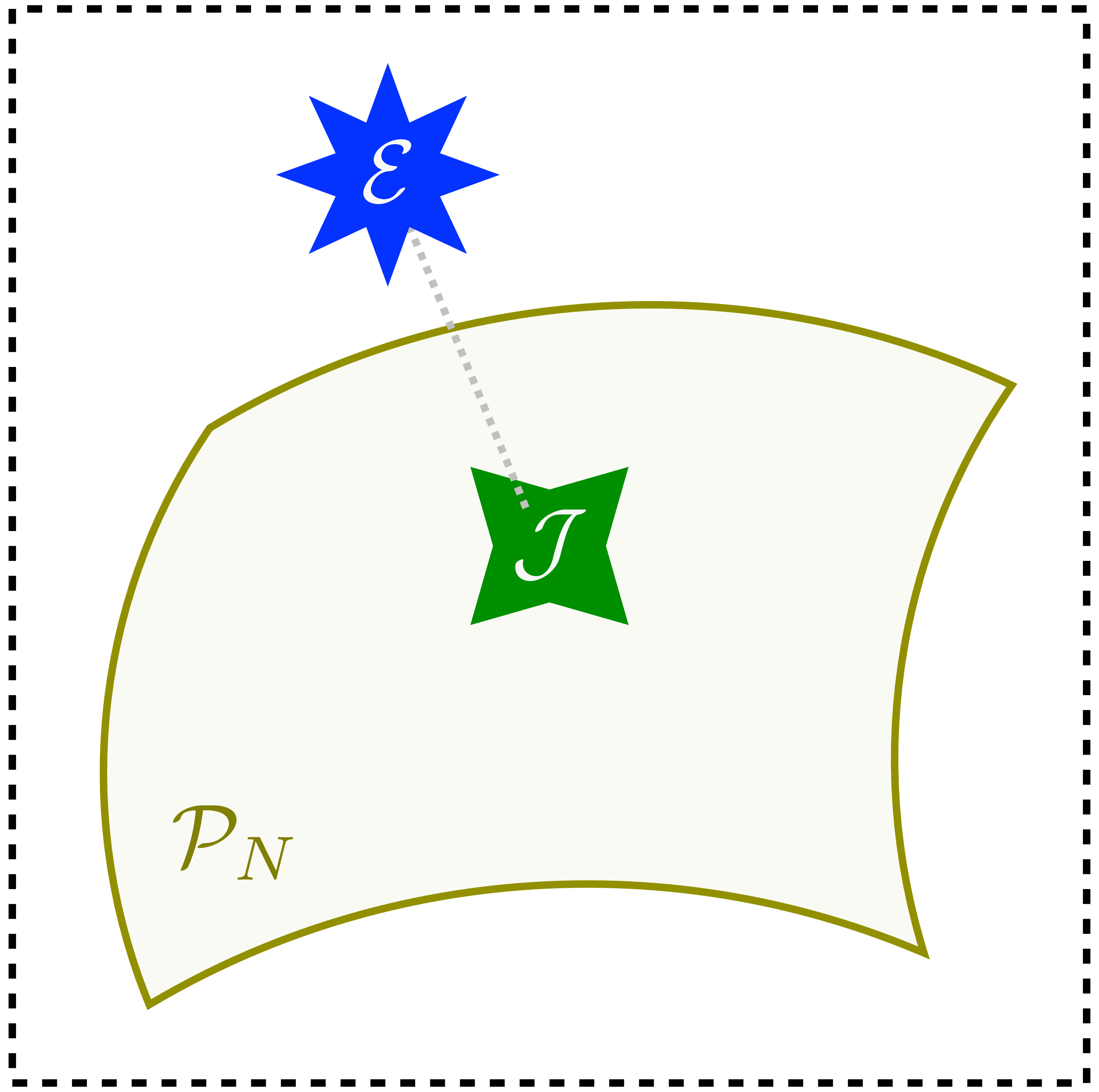
[PTK, Metodiev, Thaler, 2004.04159]

Jets in the Space of Events – The Closest *N*-particle Description of an *M*-particle Event

[PTK, Metodiev, Thaler, 2004.04159]

Exclusive cone finding

XCone finds N jets by minimizing N-jettiness



$$\mathcal{J}^{\text{XCone}}_{N,\beta,R}(\mathcal{E}) = \underset{\mathcal{J}\in\mathcal{P}_N}{\arg\min}\, \text{EMD}_{\beta,R}(\mathcal{E},\mathcal{J})$$

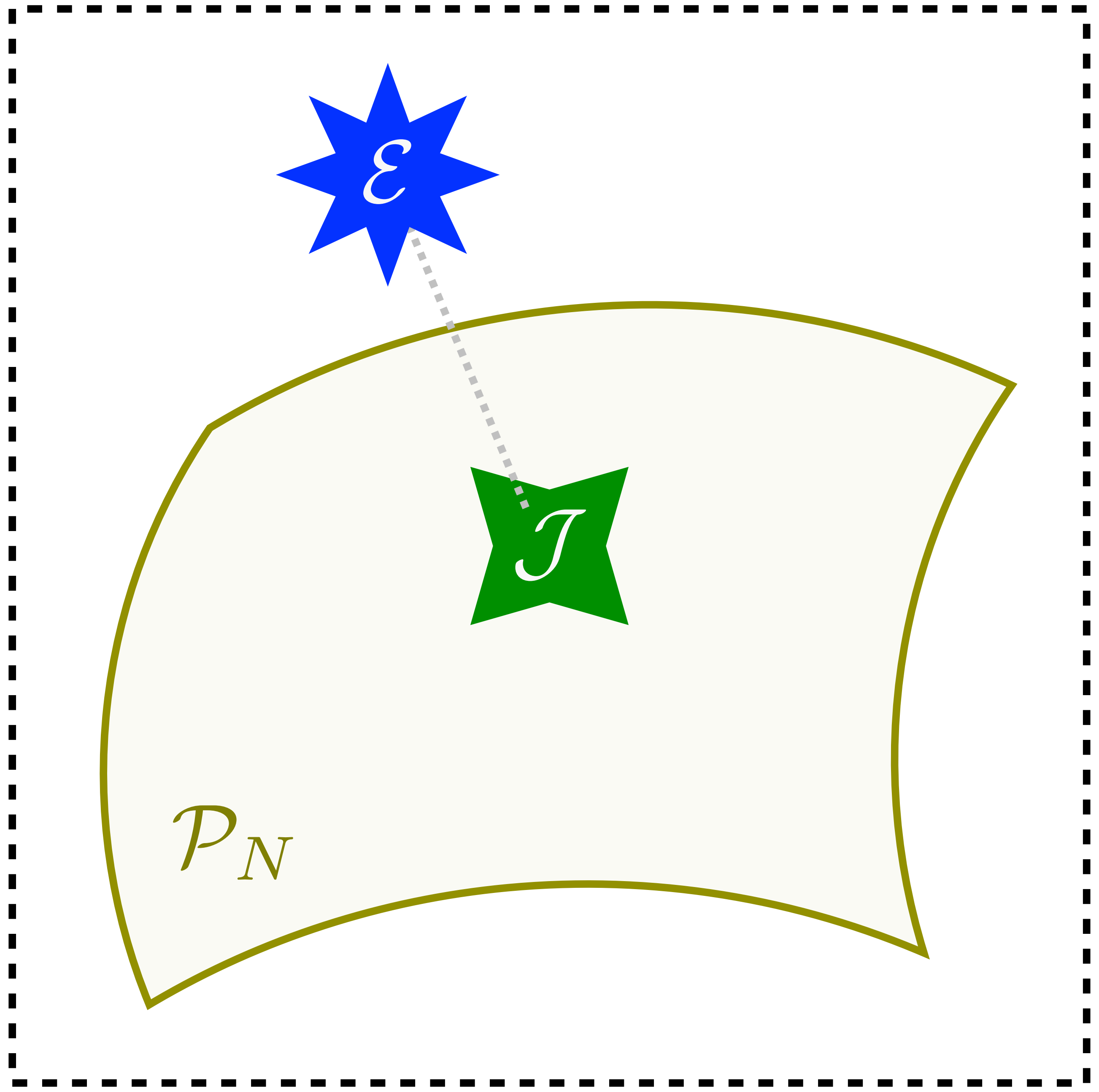
[Stewart, Tackmann, Thaler, Vermilion, Wilkason, JHEP 2015;
Thaler, Wilkason, JHEP 2015]

Jets in the Space of Events – The Closest N-particle Description of an M-particle Event

[PTK, Metodiev, Thaler, 2004.04159]

Exclusive cone finding

XCone finds N jets by minimizing N-jettiness

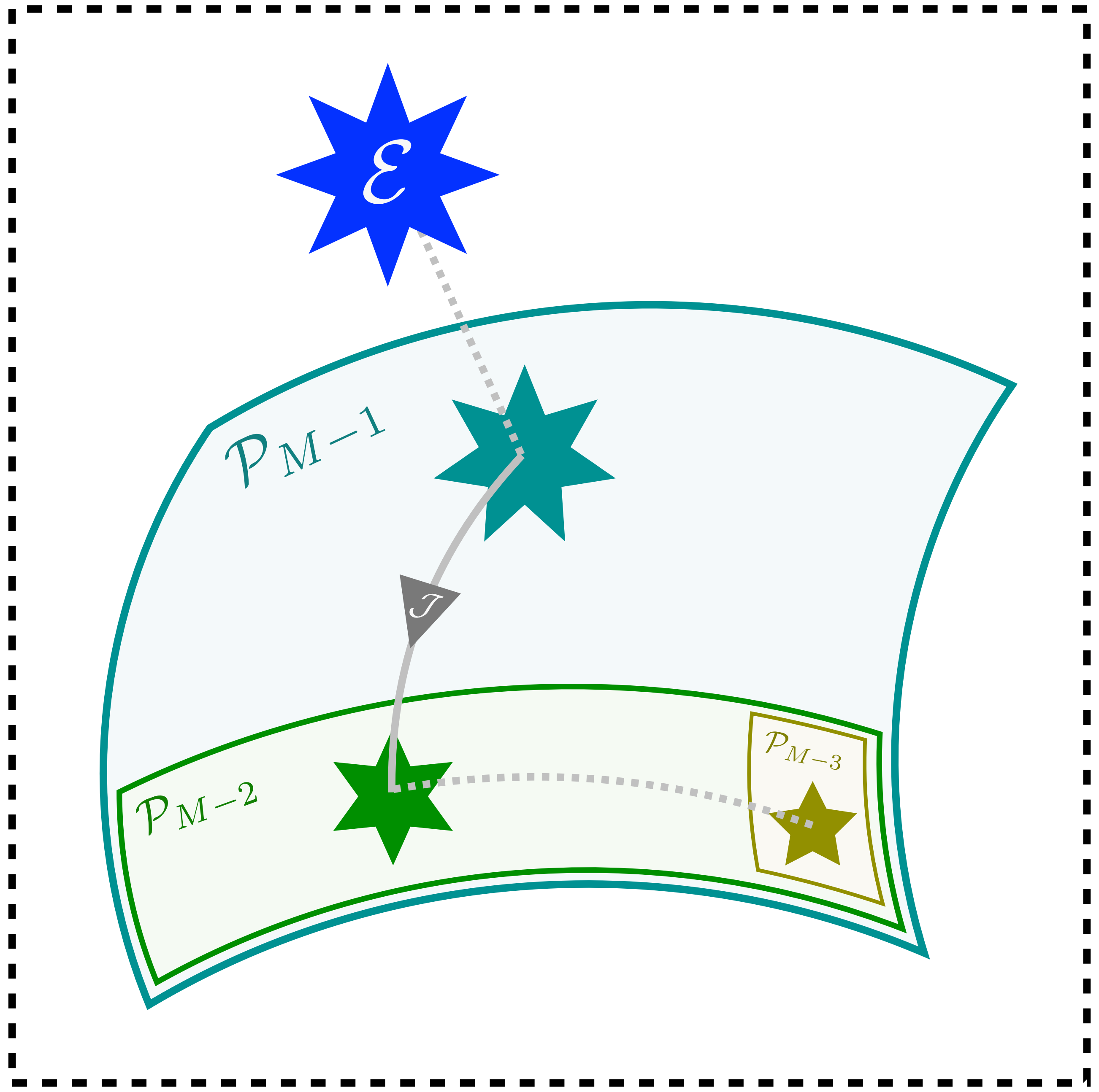


$$\mathcal{J}^{\text{XCone}}_{N,\beta,R}(\mathcal{E}) = \underset{\mathcal{J}\in\mathcal{P}_N}{\arg\min}\, \text{EMD}_{\beta,R}(\mathcal{E},\mathcal{J})$$

[Stewart, Tackmann, Thaler, Vermilion, Wilkason, JHEP 2015; Thaler, Wilkason, JHEP 2015]

Sequential recombination

Iteratively merges particles or identifies a jet



"destroying" energy corresponds to identifying a jet

event with one fewer particle after one step

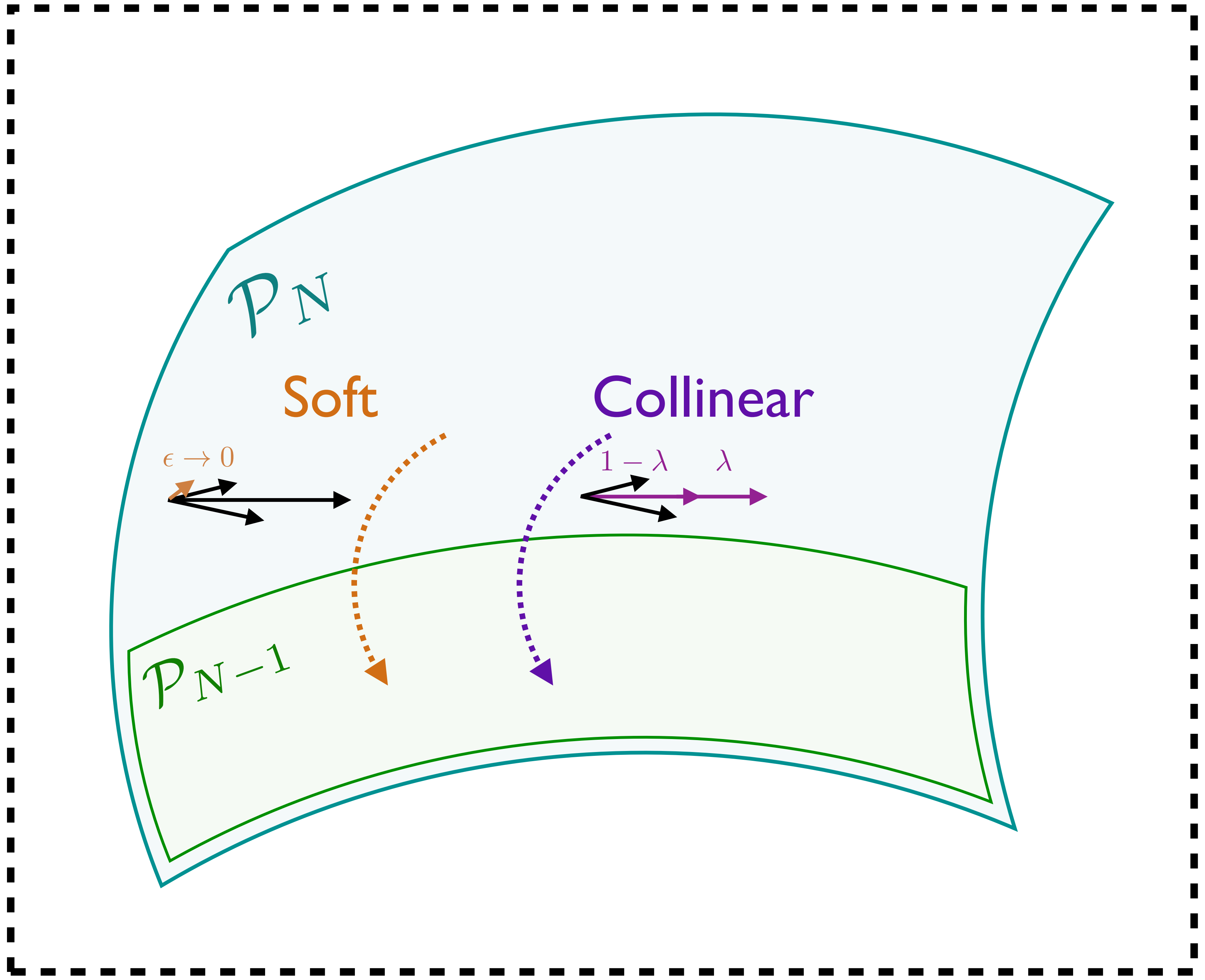
$$\mathcal{E}^{(\beta,R)}_{M-1}(\mathcal{E}_M) = \underset{\mathcal{E}'_{M-1}\in\mathcal{P}_{M-1}}{\arg\min}\, \text{EMD}_{\beta,R}(\mathcal{E}_M,\mathcal{E}'_{M-1})$$

[Catani, Dokshitzer, Seymour, Webber, Nucl. Phys. B 1993; Ellis, Soper, PRD 1993; Dokshitzer, Leder, Moretti, Webber, JHEP 1997; Cacciari, Salam, Soyez, JHEP 2008]

Perturbation Theory in the Space of Events

[PTK, Metodiev, Thaler, 2004.04159]

Infrared singularities of massless gauge theories appear on each $\mathcal{P}_N$



Perturbation Theory in the Space of Events

[PTK, Metodiev, Thaler, 2004.04159]

Sudakov safety

[Larkoski, Thaler, JHEP 2014; Larkoski, Marzani, Thaler, PRD 2015]

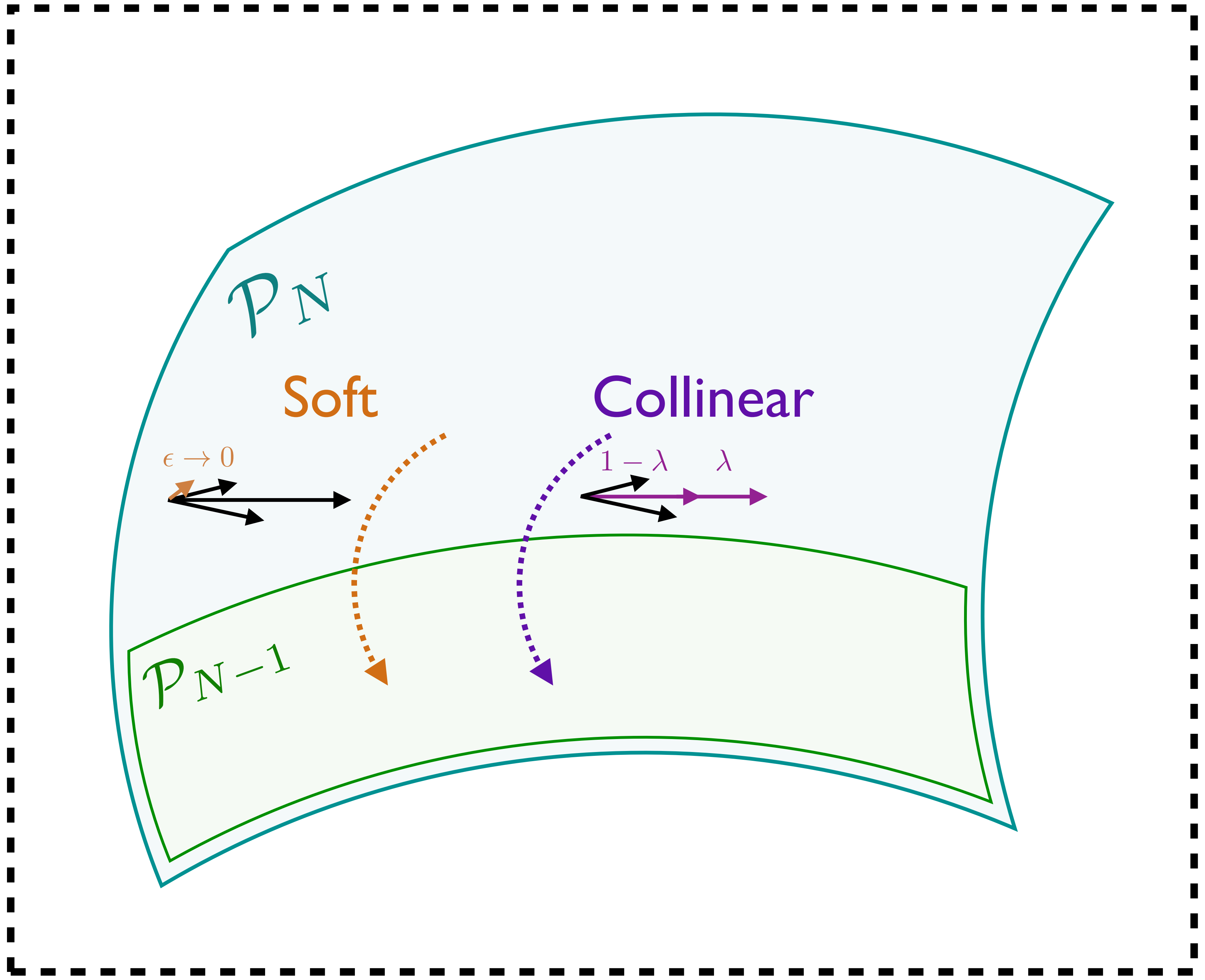
Some observables have discontinuities on P_N for some N

A resummed IRC-safe companion can mitigate the divergences

$$p(\mathcal{O}_{\text{Sudakov}}) = \int d\mathcal{O}_{\text{Comp.}}\, p(\mathcal{O}_{\text{Sudakov}}|\mathcal{O}_{\text{Comp.}})\, p(\mathcal{O}_{\text{Comp.}})$$

Event geometry suggests N-(sub)jettiness as universal companion

Infrared singularities of massless gauge theories appear on each P_N



Perturbation Theory in the Space of Events

[PTK, Metodiev, Thaler, 2004.04159]

Sudakov safety

[Larkoski, Thaler, JHEP 2014; Larkoski, Marzani, Thaler, PRD 2015]

Some observables have discontinuities on P_N for some N

A resummed IRC-safe companion can mitigate the divergences

$$p(\mathcal{O}_{\text{Sudakov}}) = \int d\mathcal{O}_{\text{Comp.}}\, p(\mathcal{O}_{\text{Sudakov}} | \mathcal{O}_{\text{Comp.}})\, p(\mathcal{O}_{\text{Comp.}})$$

Event geometry suggests N-(sub)jettiness as universal companion

Fixed-order calculability

[Sterman, PRD 1979; Banfi, Salam, Zanderighi, JHEP 2005]

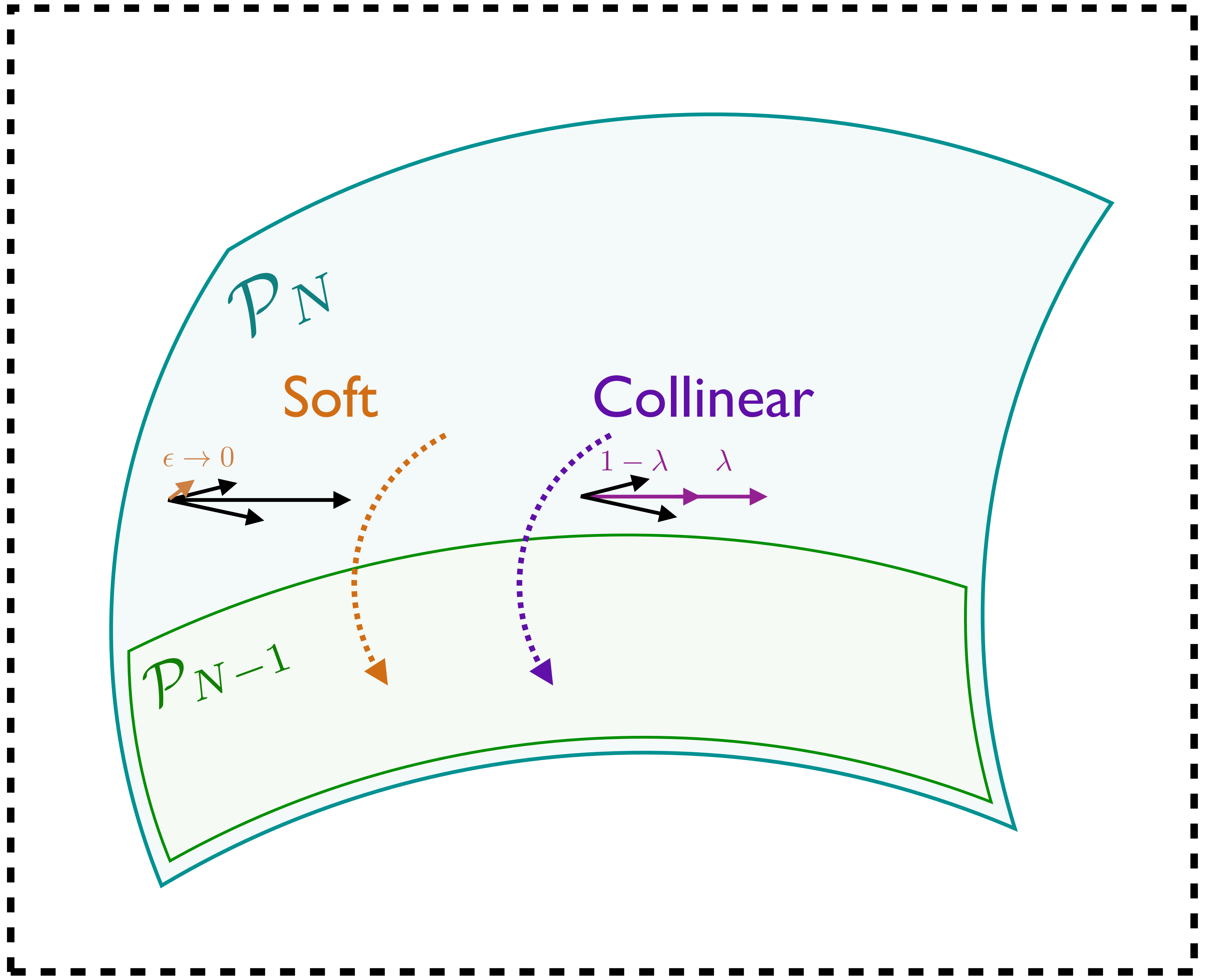
Is a statement of integrability on each P_N

EMD continuity must be upgraded to EMD-Hölder continuity on each P_N

$$\lim_{\mathcal{E} \to \mathcal{E}'} \frac{\mathcal{O}(\mathcal{E}) - \mathcal{O}(\mathcal{E}')}{\text{EMD}(\mathcal{E}, \mathcal{E}')^c} = 0, \quad c > 0$$

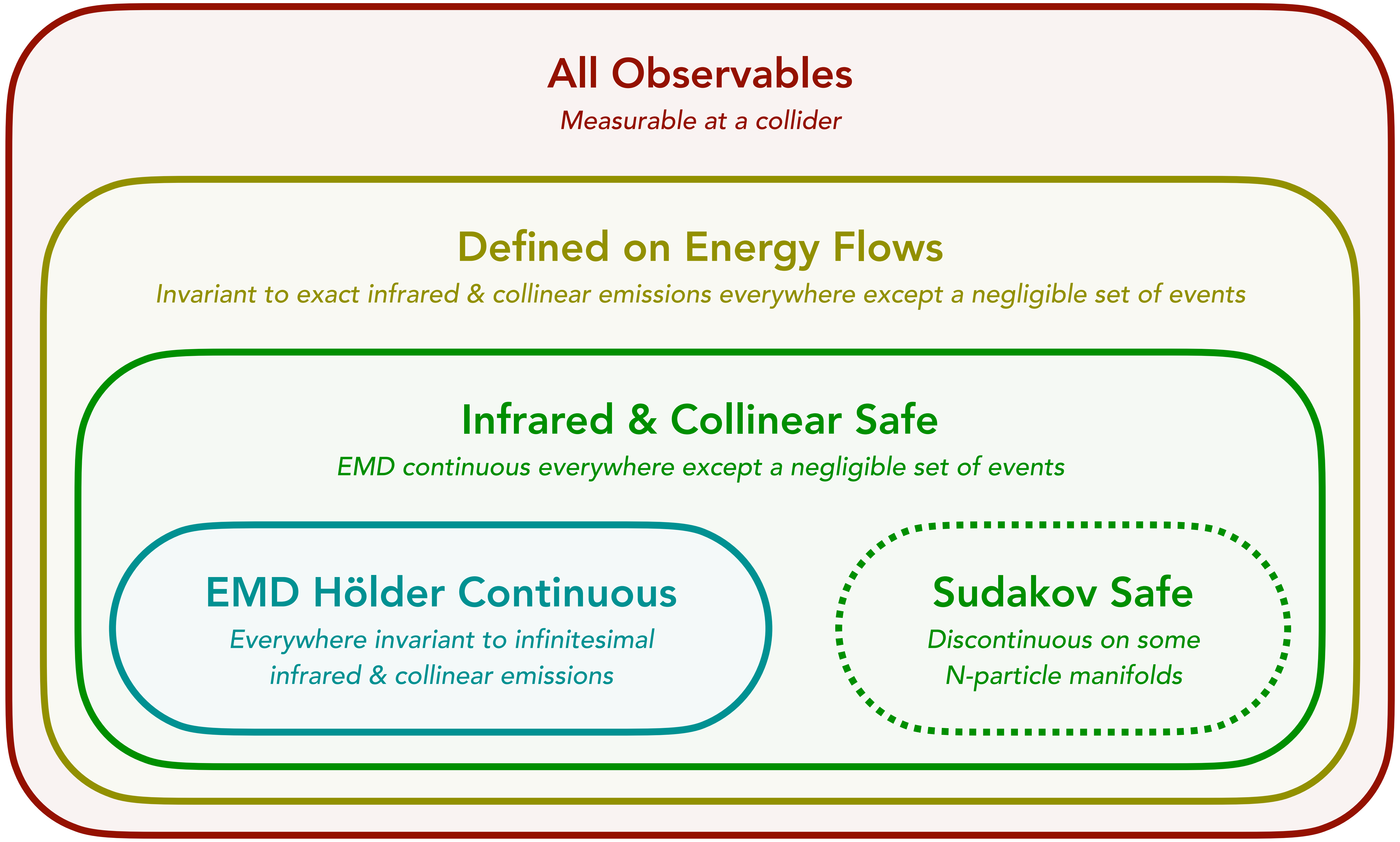
Example: $V(\mathcal{E}) = \mathcal{T}_2(\mathcal{E}) \left(1 + \frac{1}{\ln E(\mathcal{E})/\mathcal{T}_3(\mathcal{E})}\right)$ is EMD continuous but not EMD Hölder continuous (it is Sudakov safe)

Infrared singularities of massless gauge theories appear on each P_N



Hierarchy of IRC Safety Definitions

[PTK, Metodiev, Thaler, 2004.04159]



All Observables	Comments
Multiplicity ($\sum_i 1$)	IR unsafe and C unsafe
Momentum Dispersion [65] ($\sum_i E_i^2$)	IR safe but C unsafe
Sphericity Tensor [66] ($\sum_i p_i^\mu p_i^\nu$)	IR safe but C unsafe
Number of Non-Zero Calorimeter Deposits	C safe but IR unsafe
Defined on Energy Flows	
Pseudo-Multiplicity ($\min\{N \mid \mathcal{T}_N = 0\}$)	Robust to exact IR or C emissions
Infrared & Collinear Safe	
Jet Energy ($\sum_i E_i$)	Disc. at jet boundary
Heavy Jet Mass [67]	Disc. at hemisphere boundary
Soft-Dropped Jet Mass [38, 68]	Disc. at grooming threshold
Calorimeter Activity [69] (N_{95})	Disc. at cell boundary
Sudakov Safe	
Groomed Momentum Fraction [39] (z_g)	Disc. on 1-particle manifold
Jet Angularity Ratios [37]	Disc. on 1-particle manifold
N-subjettiness Ratios [47, 48] (τ_{N+1}/τ_N)	Disc. on N-particle manifold
V parameter [36] (Eq. (2.11))	Hölder disc. on 3-particle manifold
EMD Hölder Continuous Everywhere	
Thrust [40, 41]	
Spherocity [42]	
Angularities [70]	
N-jettiness [44] ($\mathcal{T}_N$)	
C parameter [71–74]	Resummation beneficial at $C = \frac{3}{4}$
Linear Sphericity [72] ($\sum_i E_i n_i^\mu n_i^\nu$)	
Energy Correlators [36, 75–77]	
Energy Flow Polynomials [15, 17]	

Six Decades of Collider Techniques as EMD Geometry

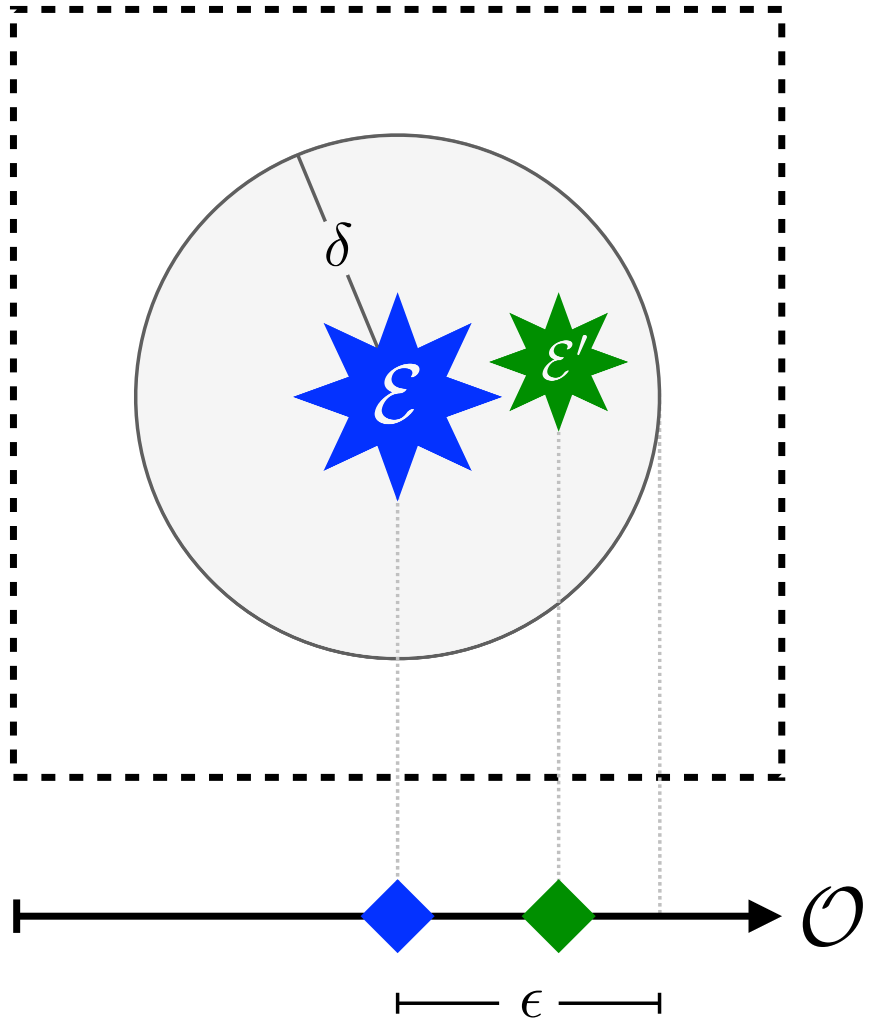
[PTK, Metodiev, Thaler, JHEP 2020]

(Timeline design by Eric Metodiev)

Six Decades of Collider Techniques as EMD Geometry

[PTK, Metodiev, Thaler, JHEP 2020]

IRC-safety is continuity in the space of events



Curing Infinities

1962 – 1964

Infrared Safety

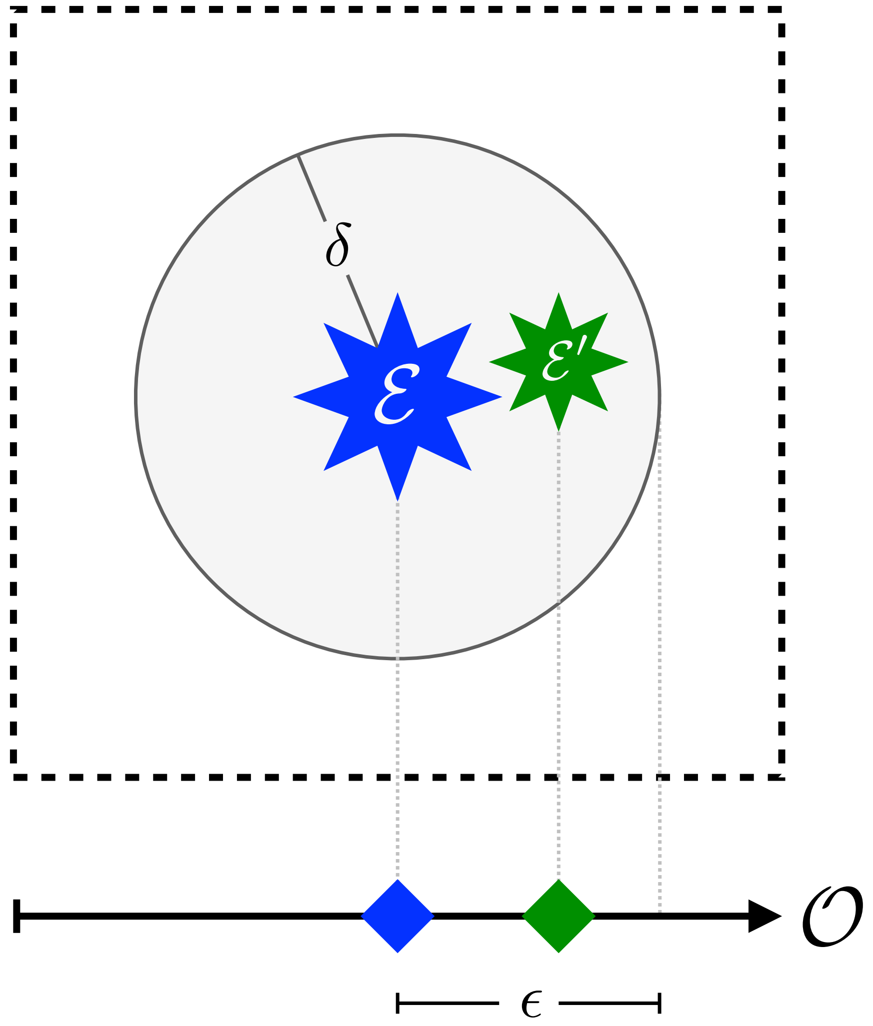
[Kinoshita, JMP 1962;
Lee, Nauenberg, PR 1964]

(Timeline design by Eric Metodiev)

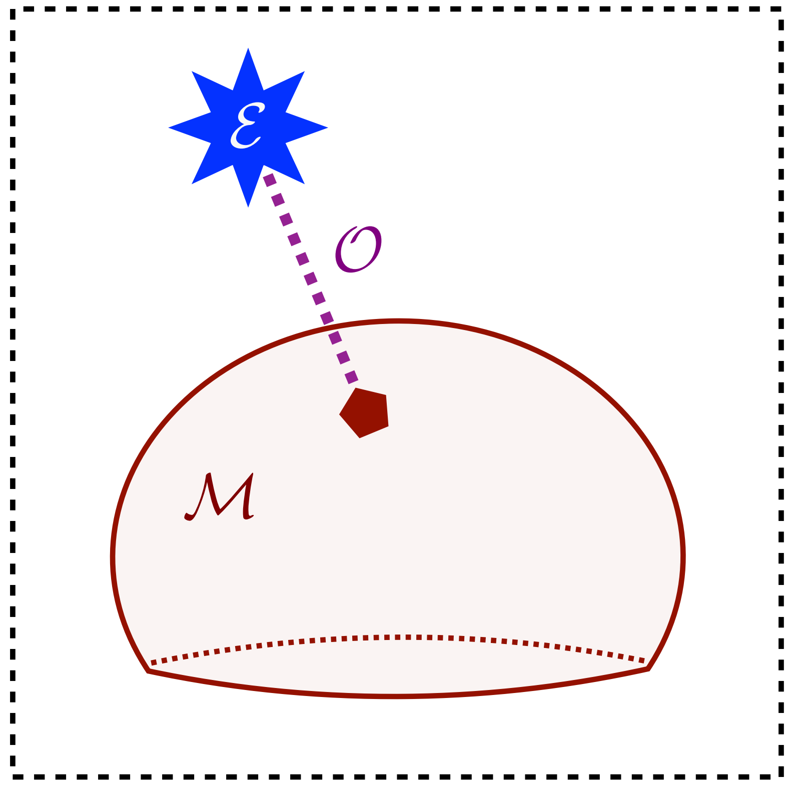
Six Decades of Collider Techniques as EMD Geometry

[PTK, Metodiev, Thaler, JHEP 2020]

IRC-safety is continuity in the space of events



Event shapes are distances from events and manifolds



$$\mathcal{O}(\mathcal{E}) = \min_{\mathcal{E}' \in \mathcal{M}} \text{EMD}_{\beta,R}(\mathcal{E}, \mathcal{E}')$$

Curing Infinities

Event Shapes

1960 — 2020

1962 – 1964
Infrared Safety
[Kinoshita, JMP 1962; Lee, Nauenberg, PR 1964]

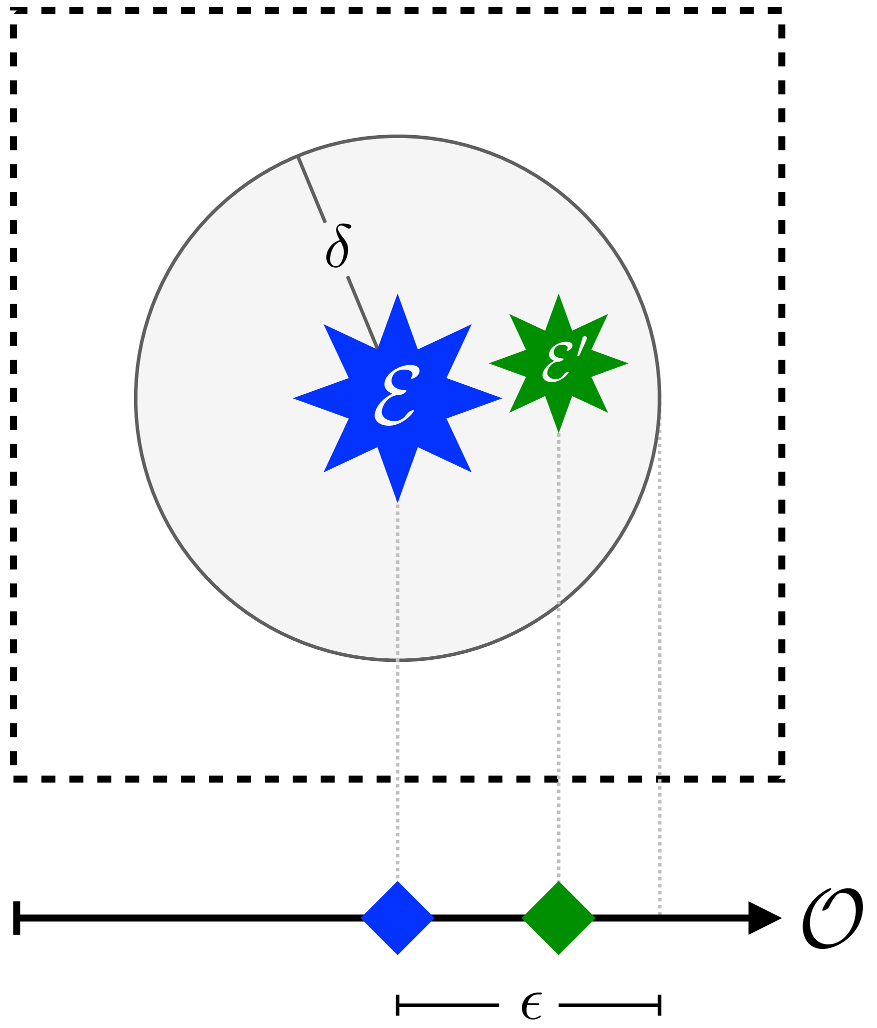
1977
Thrust, Sphericity
[Farhi, PRL 1977; Georgi, Machacek, PRL 1977]

(Timeline design by Eric Metodiev)

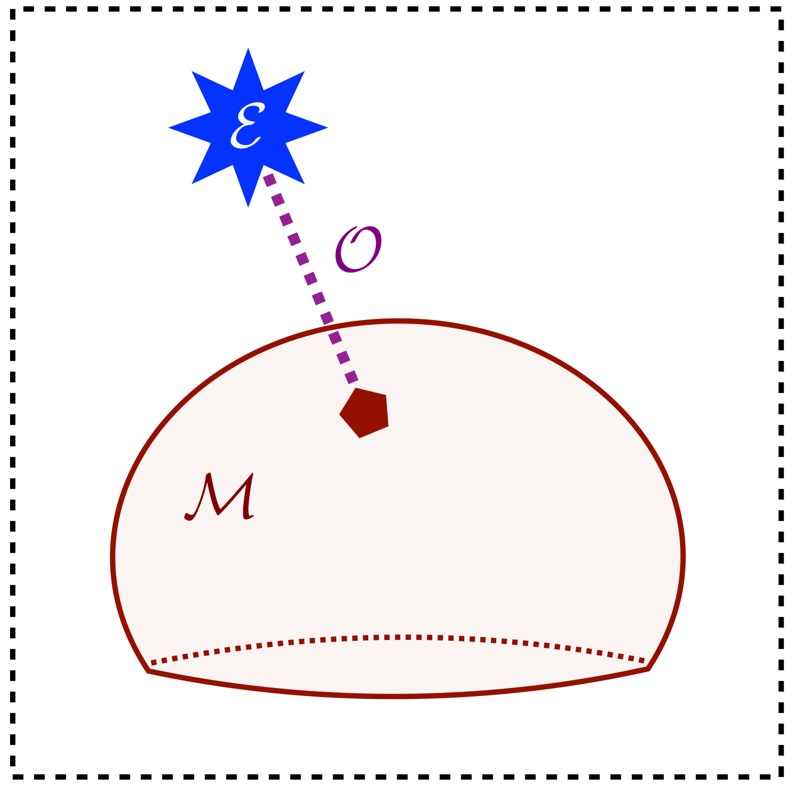
Six Decades of Collider Techniques as EMD Geometry

[PTK, Metodiev, Thaler, JHEP 2020]

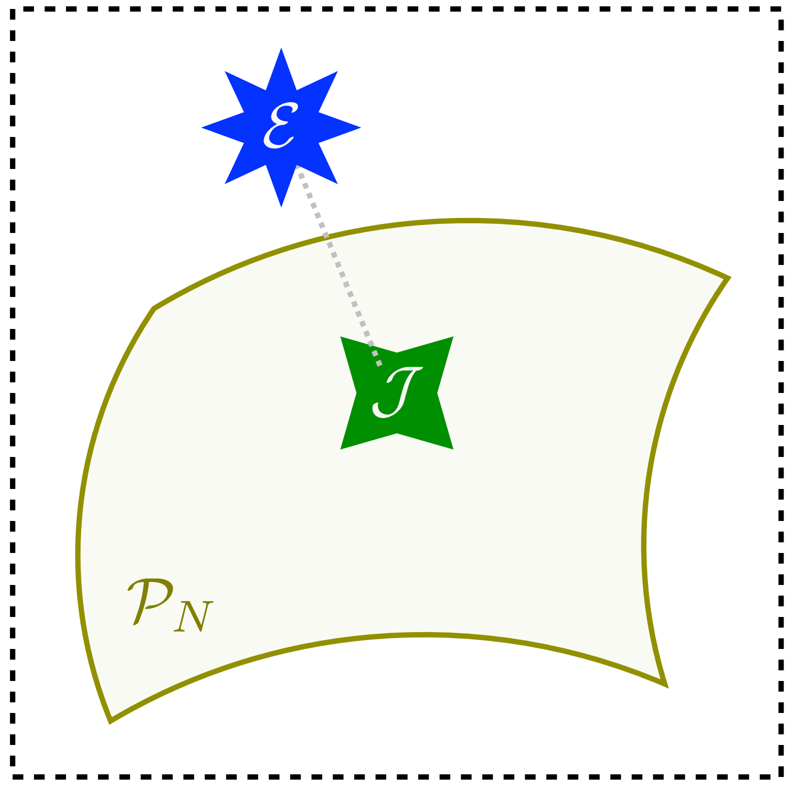
IRC-safety is continuity in the space of events



Event shapes are distances from events and manifolds



Jets are projections onto low-dimensional manifolds



$$\mathcal{O}(\mathcal{E}) = \min_{\mathcal{E}'\in\mathcal{M}} \mathrm{EMD}_{\beta,R}(\mathcal{E},\mathcal{E}')$$

$$\mathcal{J}^{\mathrm{XCone}}_{N,\beta,R}(\mathcal{E}) = \arg\min_{\mathcal{J}\in\mathcal{P}_N} \mathrm{EMD}_{\beta,R}(\mathcal{E},\mathcal{J})$$

Curing Infinities — Event Shapes — Jet Algorithms

1960 — 2020

1962 – 1964
Infrared Safety
[Kinoshita, JMP 1962;
Lee, Nauenberg, PR 1964]

1977
Thrust, Sphericity
[Farhi, PRL 1977;
Georgi, Machacek, PRL 1977]

1993
k_T Jet Clustering
[Ellis, Soper, PRD 1993;
Catani, Seymour, Dokshitzer, Webber, NPB 1993]

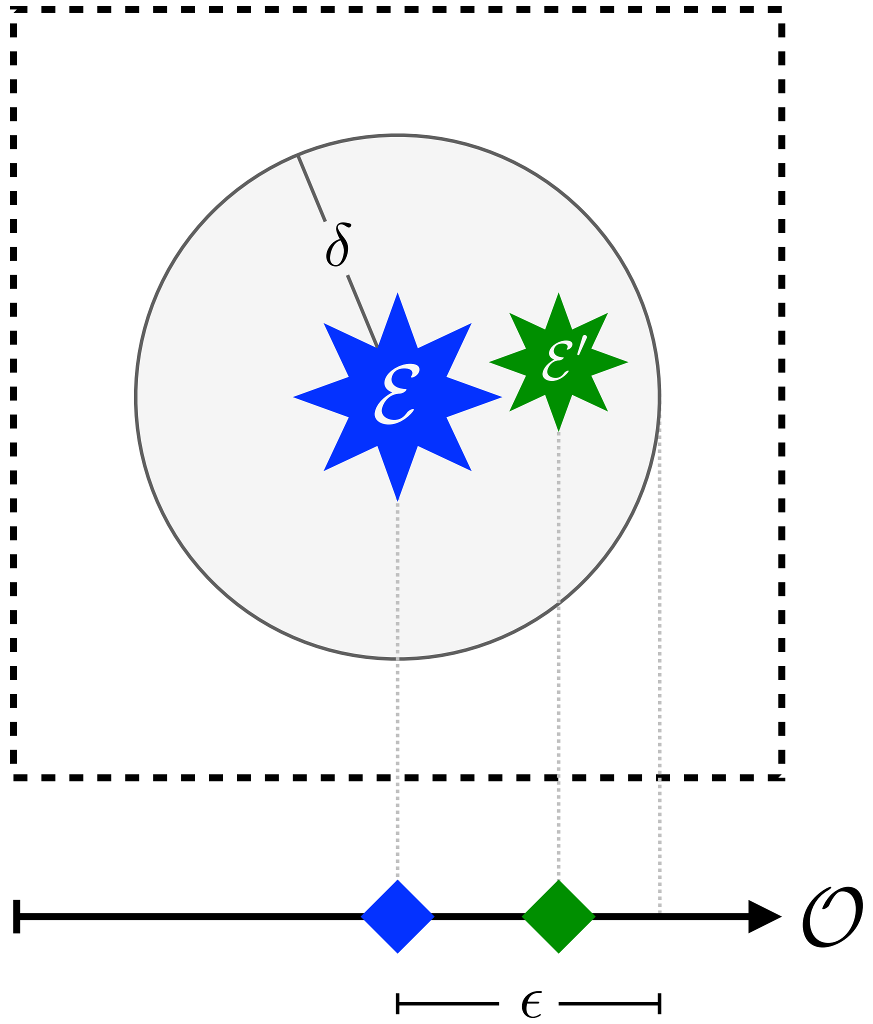
1997 – 1998
C/A Jet Clustering
[Dokshitzer, Leder, Moretti, Webber, JHEP 1997;
Wobisch, Wengler, 1998]

(Timeline design by Eric Metodiev)

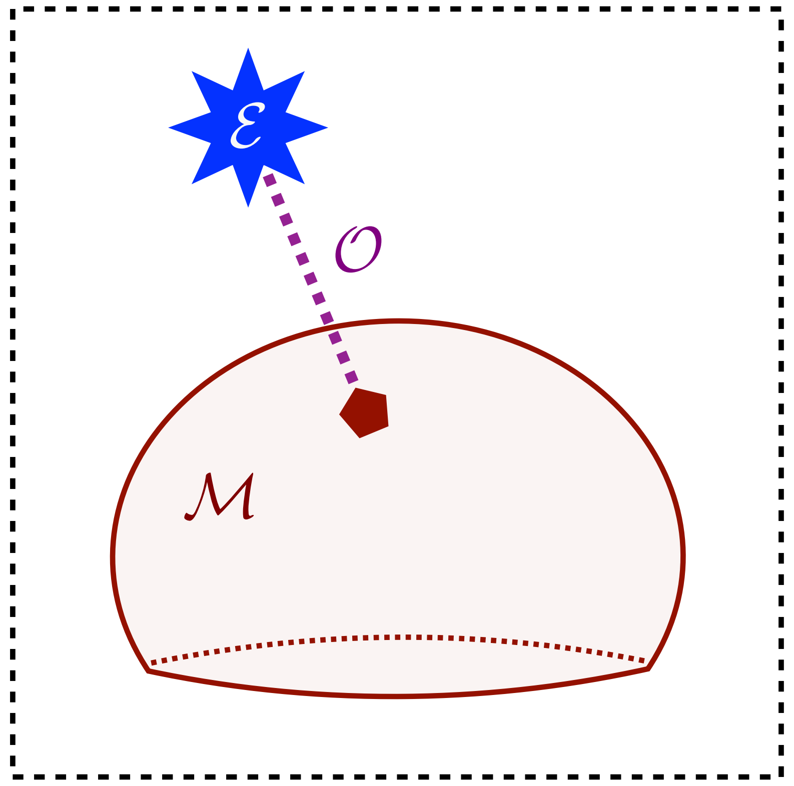
Six Decades of Collider Techniques as EMD Geometry

[PTK, Metodiev, Thaler, JHEP 2020]

IRC-safety is continuity in the space of events

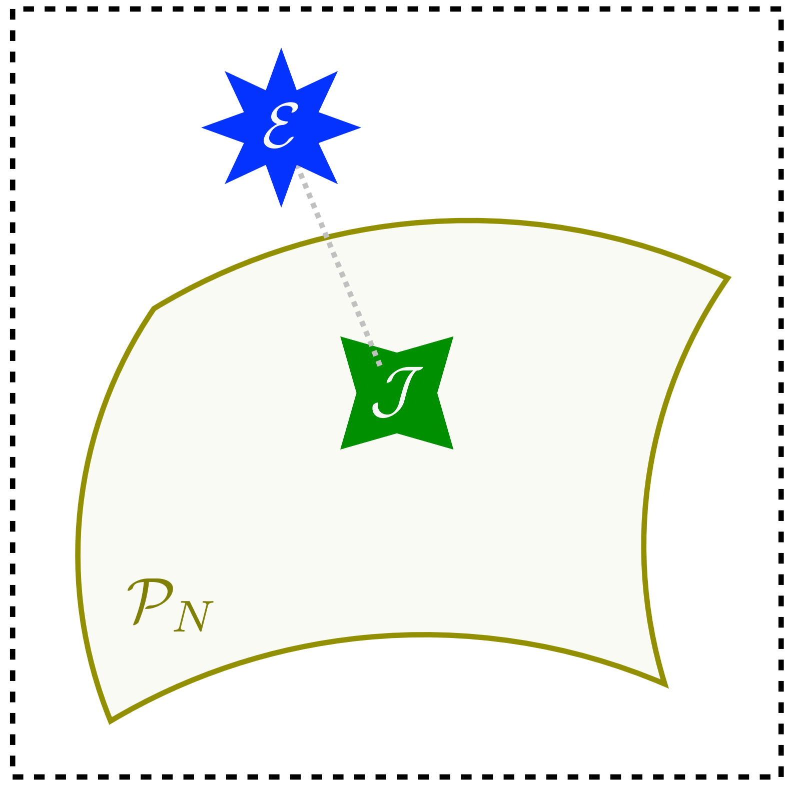


Event shapes are distances from events and manifolds



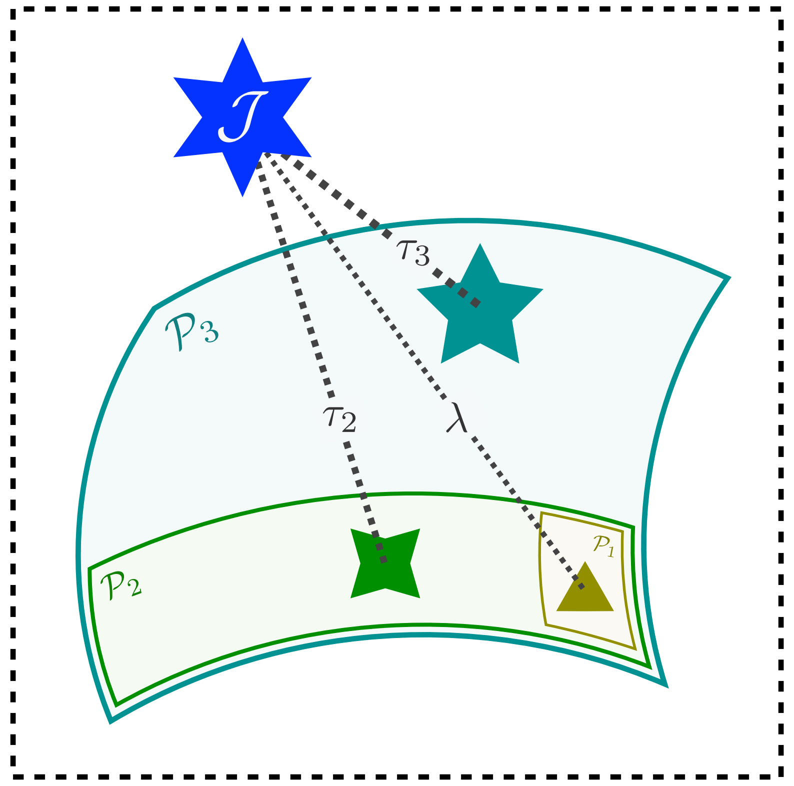
$\mathcal{O}(\mathcal{E}) = \min_{\mathcal{E}'\in\mathcal{M}} \mathrm{EMD}_{\beta,R}(\mathcal{E},\mathcal{E}')$

Jets are projections onto low-dimensional manifolds



$\mathcal{J}^{\mathrm{XCone}}_{N,\beta,R}(\mathcal{E}) = \arg\min_{\mathcal{J}\in\mathcal{P}_N} \mathrm{EMD}_{\beta,R}(\mathcal{E},\mathcal{J})$

Jet substructure probes radiation within a jet



$\tau_N^{(\beta)}(\mathcal{J}) = \min_{\mathcal{J}'\in\mathcal{P}_N} \mathrm{EMD}_{\beta}(\mathcal{J},\mathcal{J}')$

Curing Infinities

Event Shapes

Jet Algorithms

Jet Substructure

1960 — 2020

1962 – 1964
Infrared Safety
[Kinoshita, JMP 1962; Lee, Nauenberg, PR 1964]

1977
Thrust, Sphericity
[Farhi, PRL 1977; Georgi, Machacek, PRL 1977]

1993
k_T Jet Clustering
[Ellis, Soper, PRD 1993; Catani, Seymour, Dokshitzer, Webber, NPB 1993]

1997 – 1998
C/A Jet Clustering
[Dokshitzer, Leder, Moretti, Webber, JHEP 1997; Wobisch, Wengler, 1998]

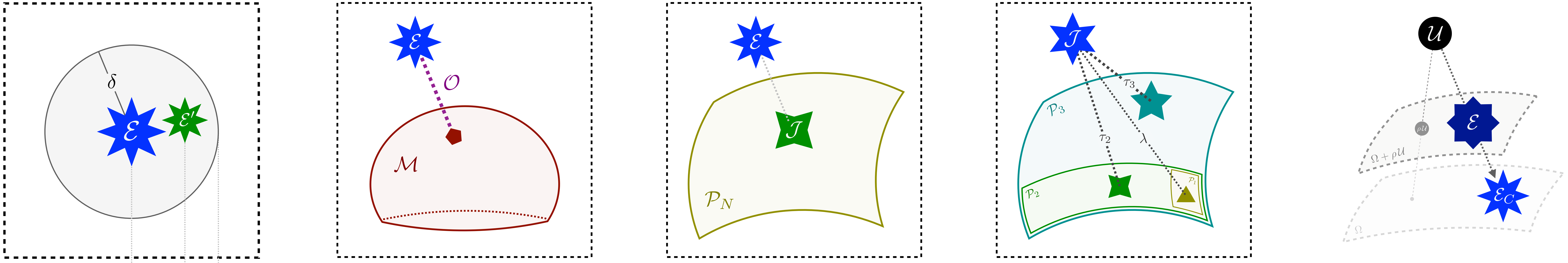
2010 – 2015
N-(sub)jetties, XCone
[Stewart, Tackmann, Waalewjin, PRL 2010; Thaler, Van Tilburg, JHEP 2011; Stewart, Tackmann, Thaler, Vermilion, Wilson, JHEP 2015]

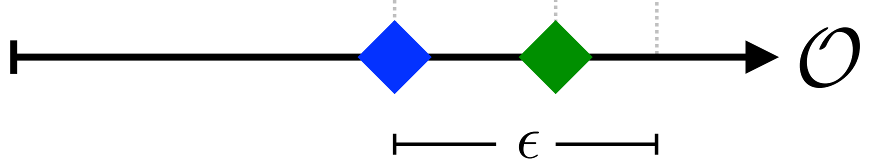
(Timeline design by Eric Metodiev)

Six Decades of Collider Techniques as EMD Geometry

[PTK, Metodiev, Thaler, JHEP 2020]

IRC-safety is continuity in the space of events	Event shapes are distances from events and manifolds	Jets are projections onto low-dimensional manifolds	Jet substructure probes radiation within a jet	Pileup mitigation moves away from the uniform event





$\mathcal{O}(\mathcal{E}) = \min_{\mathcal{E}' \in \mathcal{M}} \text{EMD}_{\beta,R}(\mathcal{E}, \mathcal{E}')$

$\mathcal{J}^{\text{XCone}}_{N,\beta,R}(\mathcal{E}) = \arg\min_{\mathcal{J} \in \mathcal{P}_N} \text{EMD}_{\beta,R}(\mathcal{E}, \mathcal{J})$

$\tau_N^{(\beta)}(\mathcal{J}) = \min_{\mathcal{J}' \in \mathcal{P}_N} \text{EMD}_\beta(\mathcal{J}, \mathcal{J}')$

$\mathcal{E}_C(\mathcal{E}, \rho) = \arg\min_{\mathcal{E}' \in \Omega} \text{EMD}_\beta(\mathcal{E}, \mathcal{E}' + \rho\mathcal{U})$

Curing Infinities — Event Shapes — Jet Algorithms — Jet Substructure — Pileup Subtraction

1960 → 2020

1962 – 1964
Infrared Safety
[Kinoshita, JMP 1962;
Lee, Nauenberg, PR 1964]

1977
Thrust, Sphericity
[Farhi, PRL 1977;
Georgi, Machacek, PRL 1977]

1993
k_T Jet Clustering
[Ellis, Soper, PRD 1993;
Catani, Seymour, Dokshitzer, Webber, NPB 1993]

1997 – 1998
C/A Jet Clustering
[Dokshitzer, Leder, Moretti, Webber, JHEP 1997;
Wobisch, Wengler, 1998]

2010 – 2015
N-(sub)jetties, XCone
[Stewart, Tackmann, Waalewjin, PRL 2010;
Thaler, Van Tilburg, JHEP 2011;
Stewart, Tackmann, Thaler, Vermilion, Wilson, JHEP 2015]

2014 – 2019
Constituent Subtraction
[Berta, Spousta, Miller, Leitner, JHEP 2014;
Berta, Masetti, Miller, Spousta, JHEP 2019]

(Timeline design by Eric Metodiev)