OmniFold

Simultaneously Unfolding All Observables

Patrick T. Komiske III

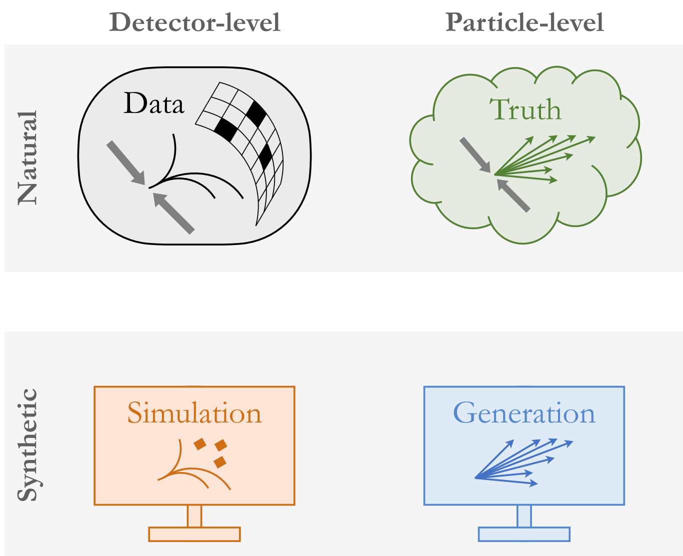
Massachusetts Institute of Technology
Center for Theoretical Physics

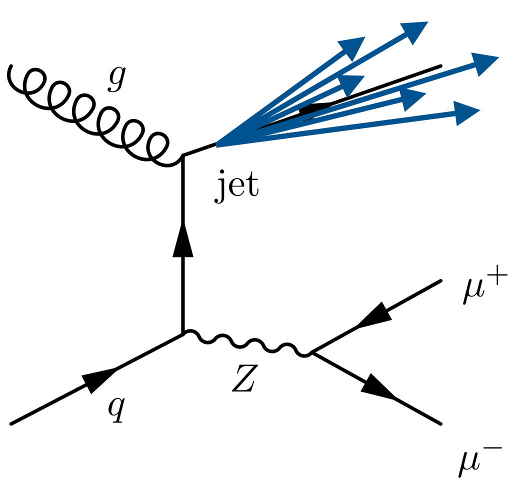
1911.09107

with Anders Andreassen, Eric Metodiev, Ben Nachman, and Jesse Thaler

ML4Jets 2020 – NYU

January 17, 2020

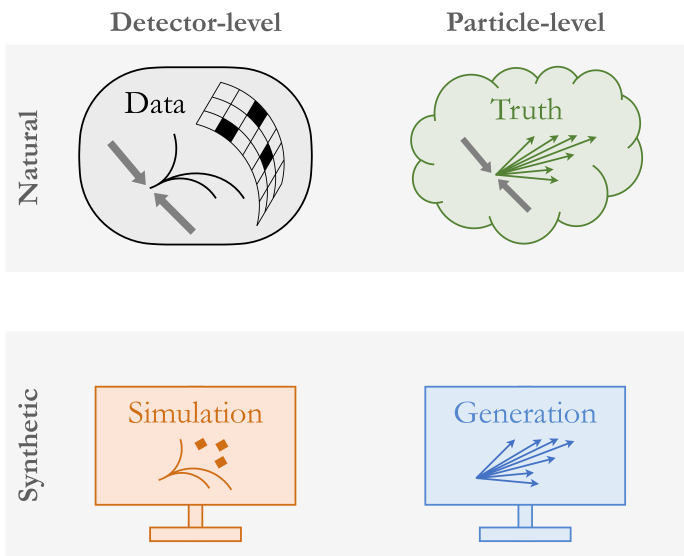


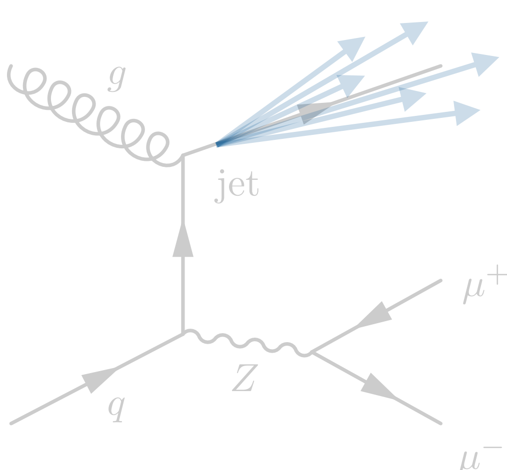


Unfolding Basics

OmniFold

Z + Jet Case Study





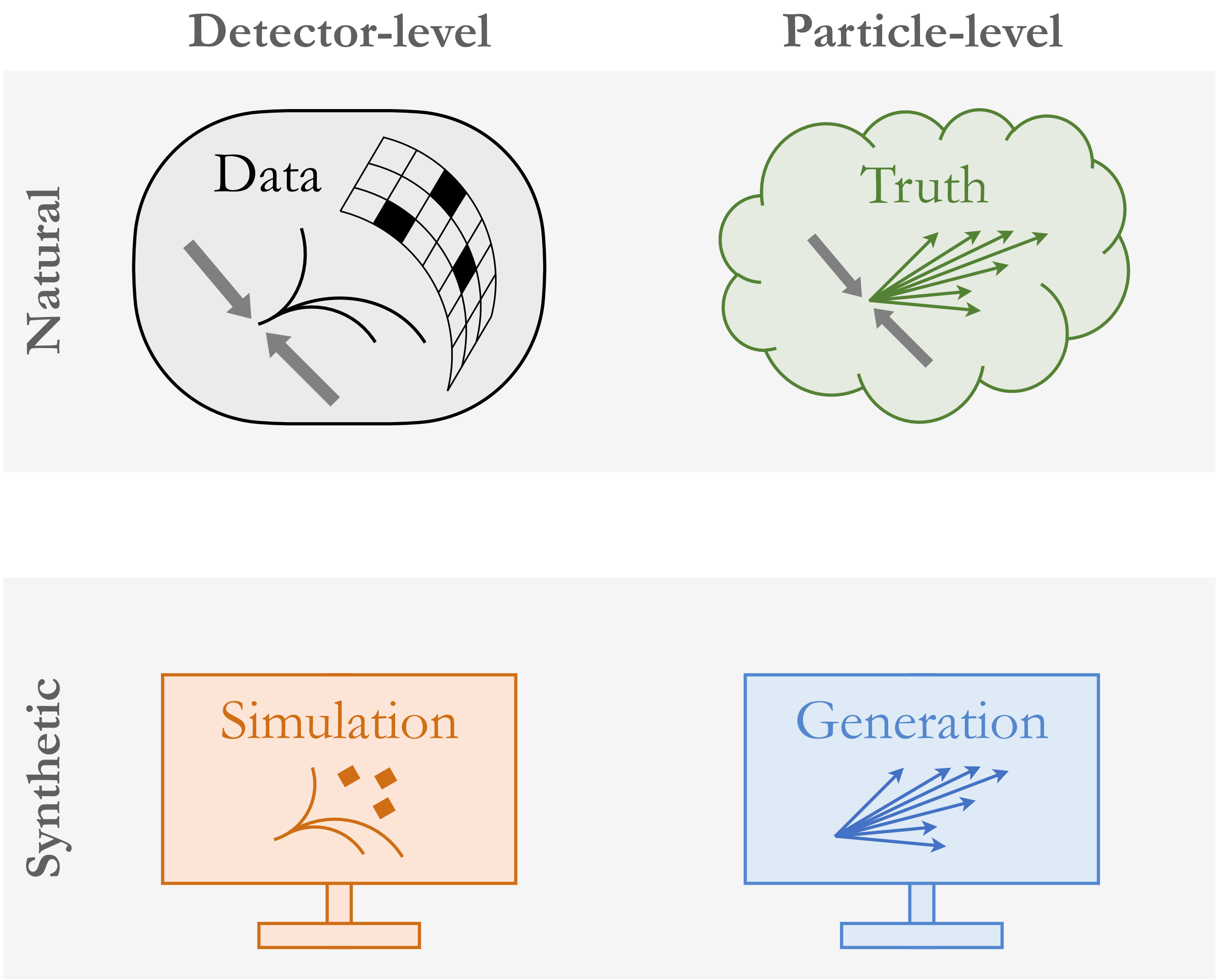
Unfolding Basics

OmniFold

Z + Jet Case Study

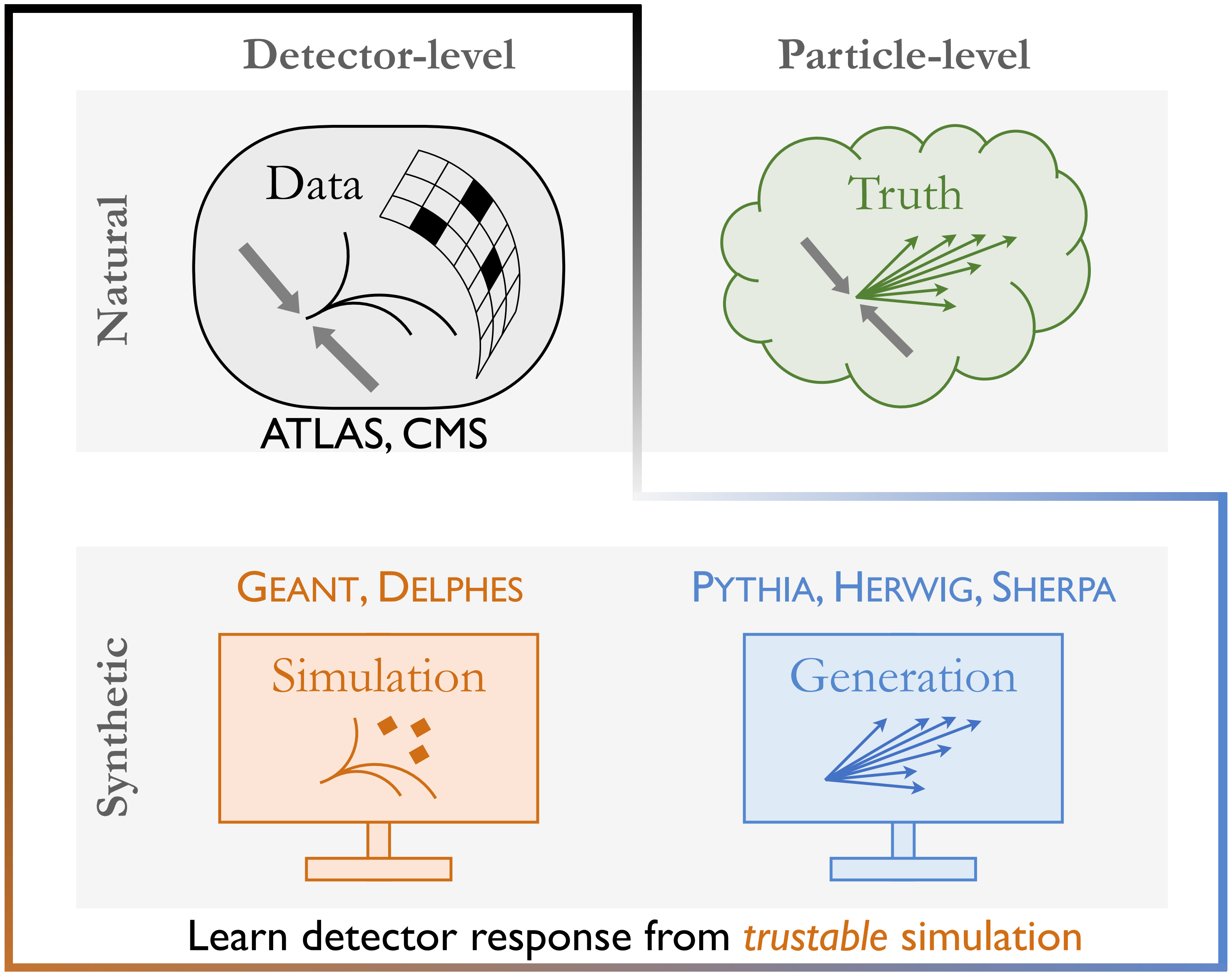
Unfolding Setup

Measurements are affected by *detector effects* of finite resolution and limited acceptance



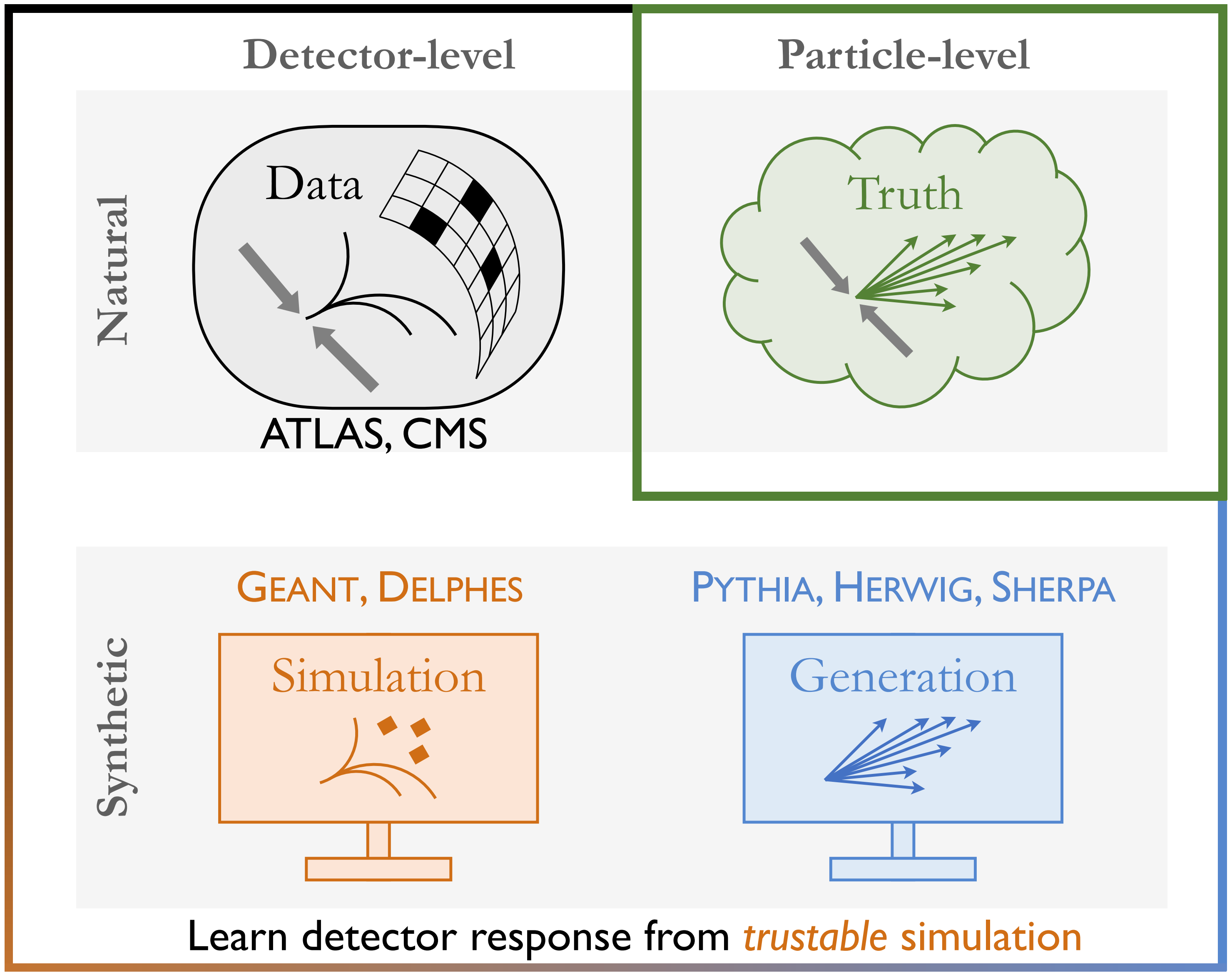
Unfolding Setup

Measurements are affected by *detector effects* of finite resolution and limited acceptance



Unfolding Setup

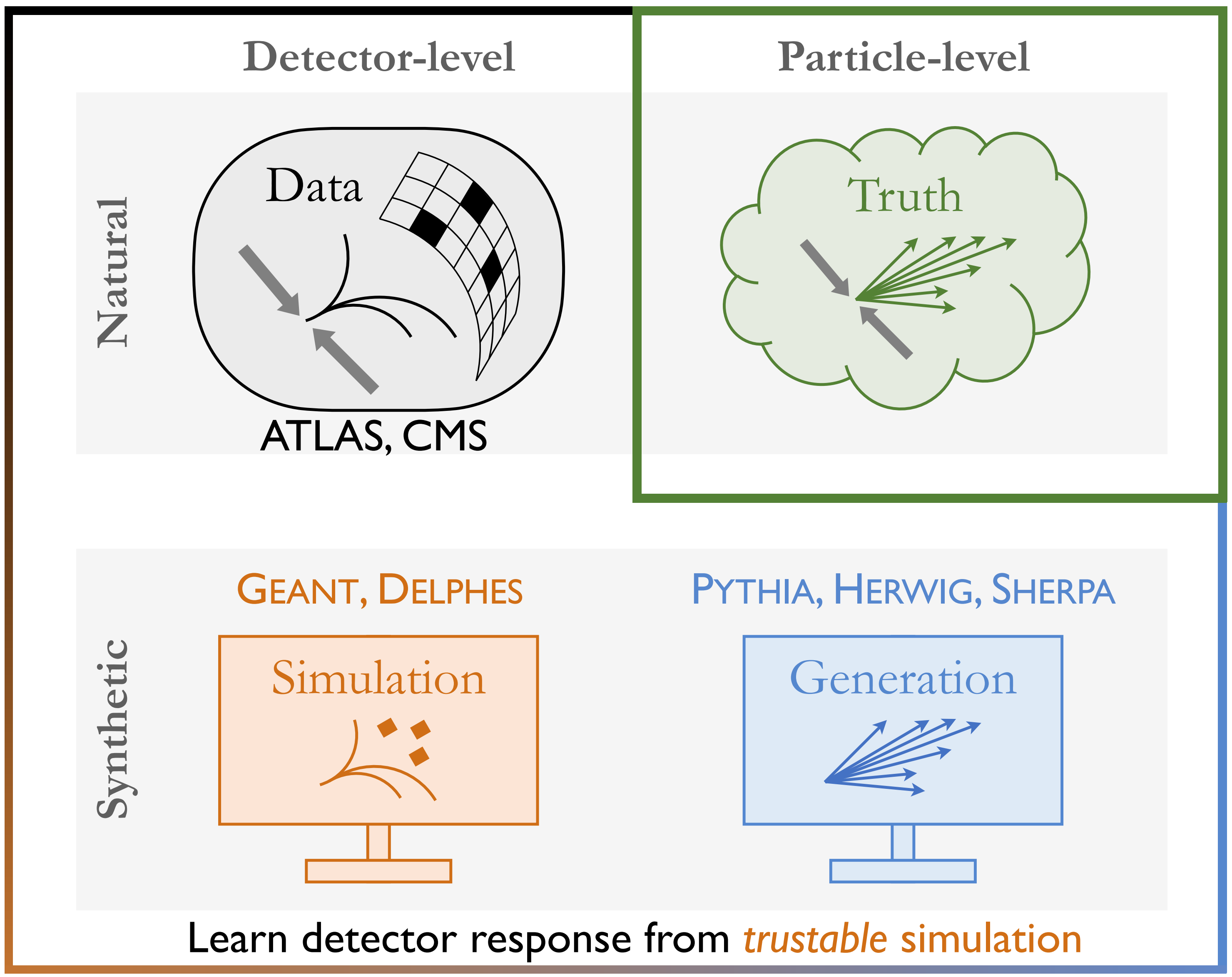
Measurements are affected by *detector effects* of finite resolution and limited acceptance



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

Unfolding Setup

Measurements are affected by *detector effects* of finite resolution and limited acceptance



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

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

Challenges with Traditional Unfolding

Previous methods are inherently binned

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

Challenges with Traditional Unfolding

Previous methods are inherently binned

Binning fixed ahead of time, cannot be changed later

Performance of method sensitive to binning

Limited number of observables

Binning induces curse of dimensionality

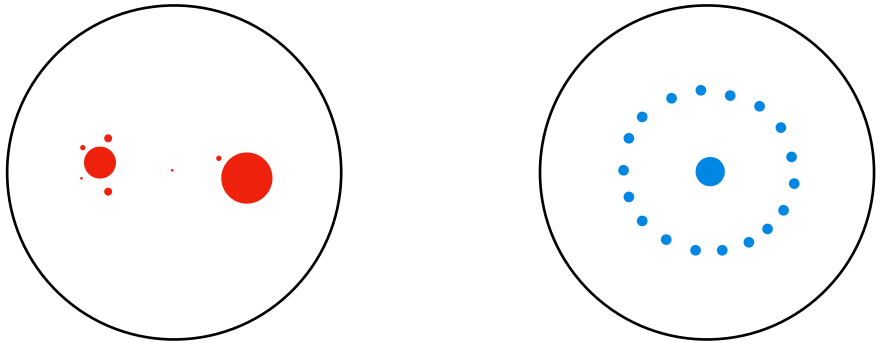
Response matrix depends on auxiliary features

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



Challenges with Traditional Unfolding

Previous methods are inherently binned

Binning fixed ahead of time, cannot be changed later

Performance of method sensitive to binning

Limited number of observables

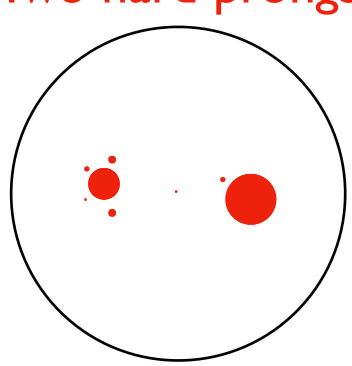
Binning induces curse of dimensionality

Response matrix depends on auxiliary features

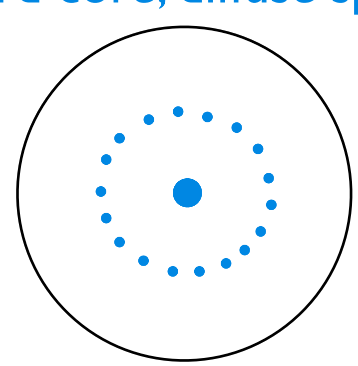
Detector-level quantity may not capture full detector effect

Example – Two jets acquiring the same mass in different ways

Jet 1
Two hard prongs

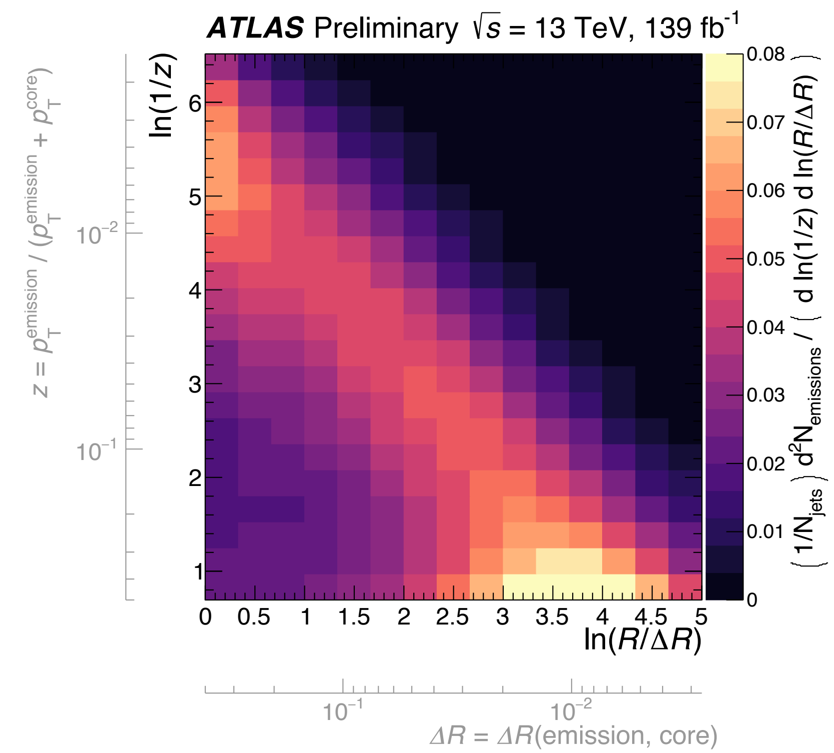


Jet 2
Hard core, diffuse spray



Example with IBU

ATLAS State-of-the-art Lund Plane Measurement
[ATLAS-CONF-2019-035]



21 x 15 bins in $\ln(1/z)$ x $\ln(R/\Delta R)$

- Must redo unfolding for other binnings e.g. finer/coarser, k_T (diagonal) binning, etc.

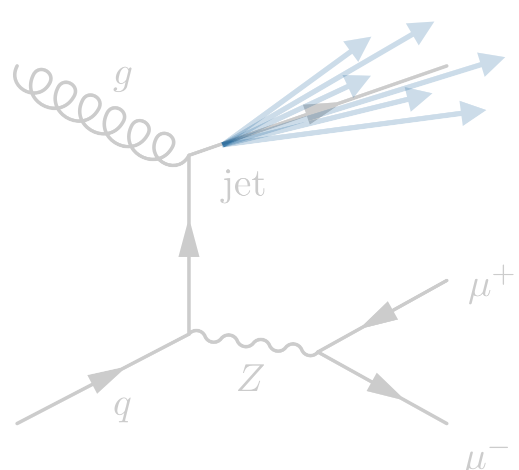
Limited to two observables

- 21^2 x 15^2 elements in response matrix R
- Going differential in n bins of p_T would multiply size of R by n^2

Unfolding Basics

OmniFold

Z + Jet Case Study



Unfolding via Likelihood Reweighting

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

Unfolding via Likelihood Reweighting

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

[Cranmer, Pavez, Louppe, 1506.02169; Andreassen, Nachman, 1907.08209]

Unfolding via Likelihood Reweighting

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

[Cranmer, Pavez, Louppe, 1506.02169; Andreassen, Nachman, 1907.08209]

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

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

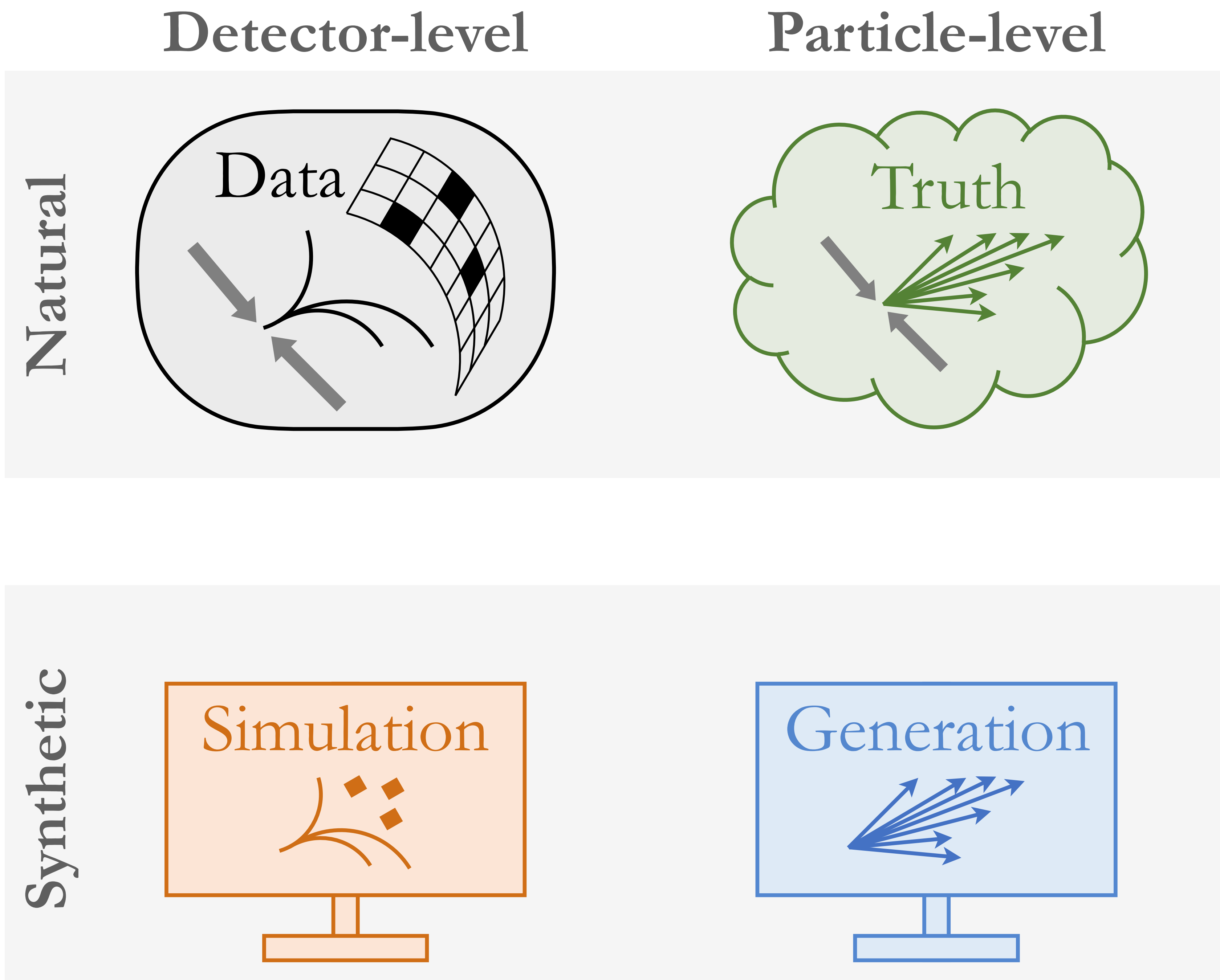
A' is statistically indistinguishable from B

Likelihood reweighting benefits from architectural improvements

OmniFold Algorithm – Schematic

[Andreassen, PTK, Metodiev, Nachman, Thaler, 1911.09107]

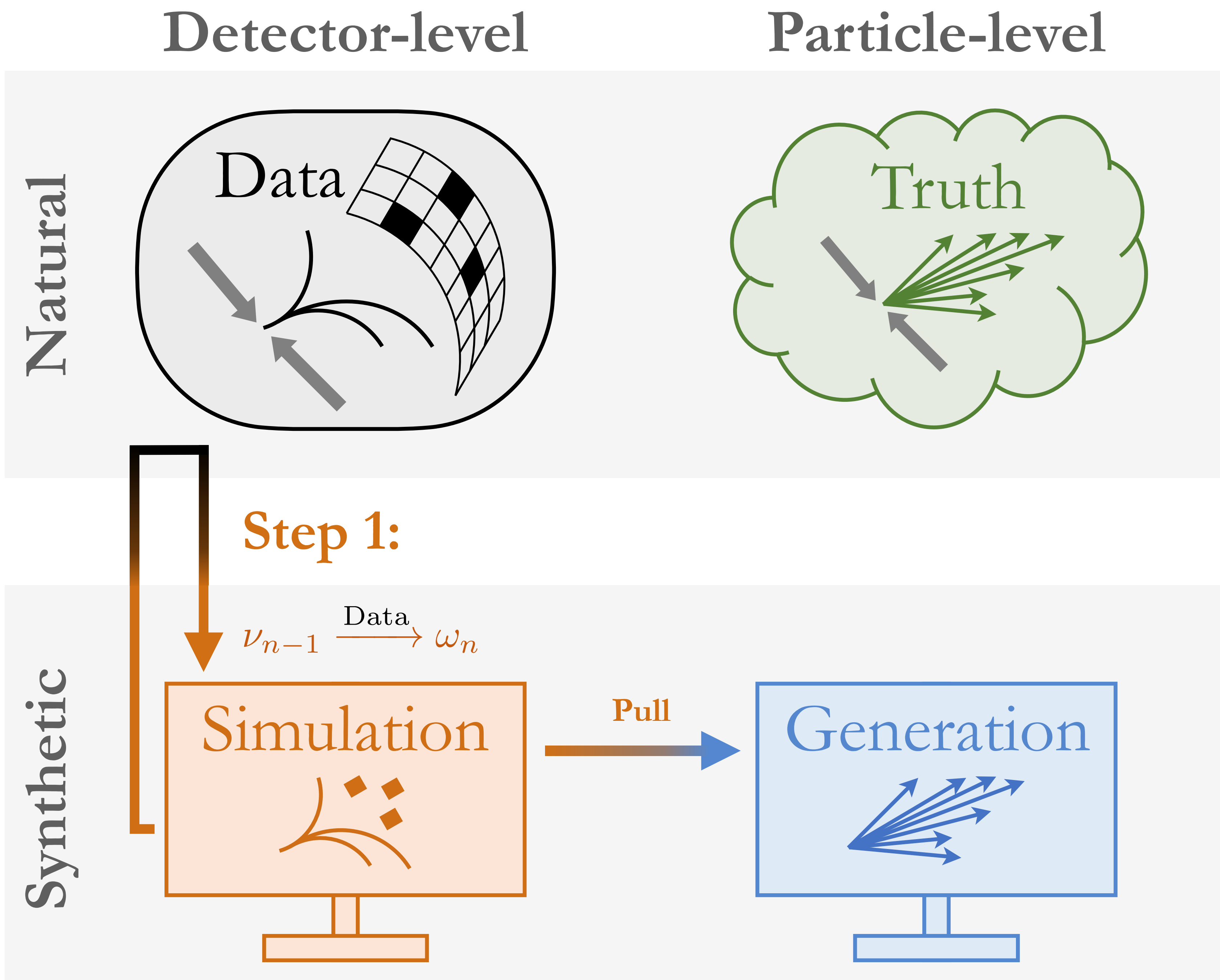
OmniFold weights particle-level Gen to be consistent with Data once passed through the detector



OmniFold Algorithm – Schematic

[Andreassen, PTK, Metodiev, Nachman, Thaler, 1911.09107]

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

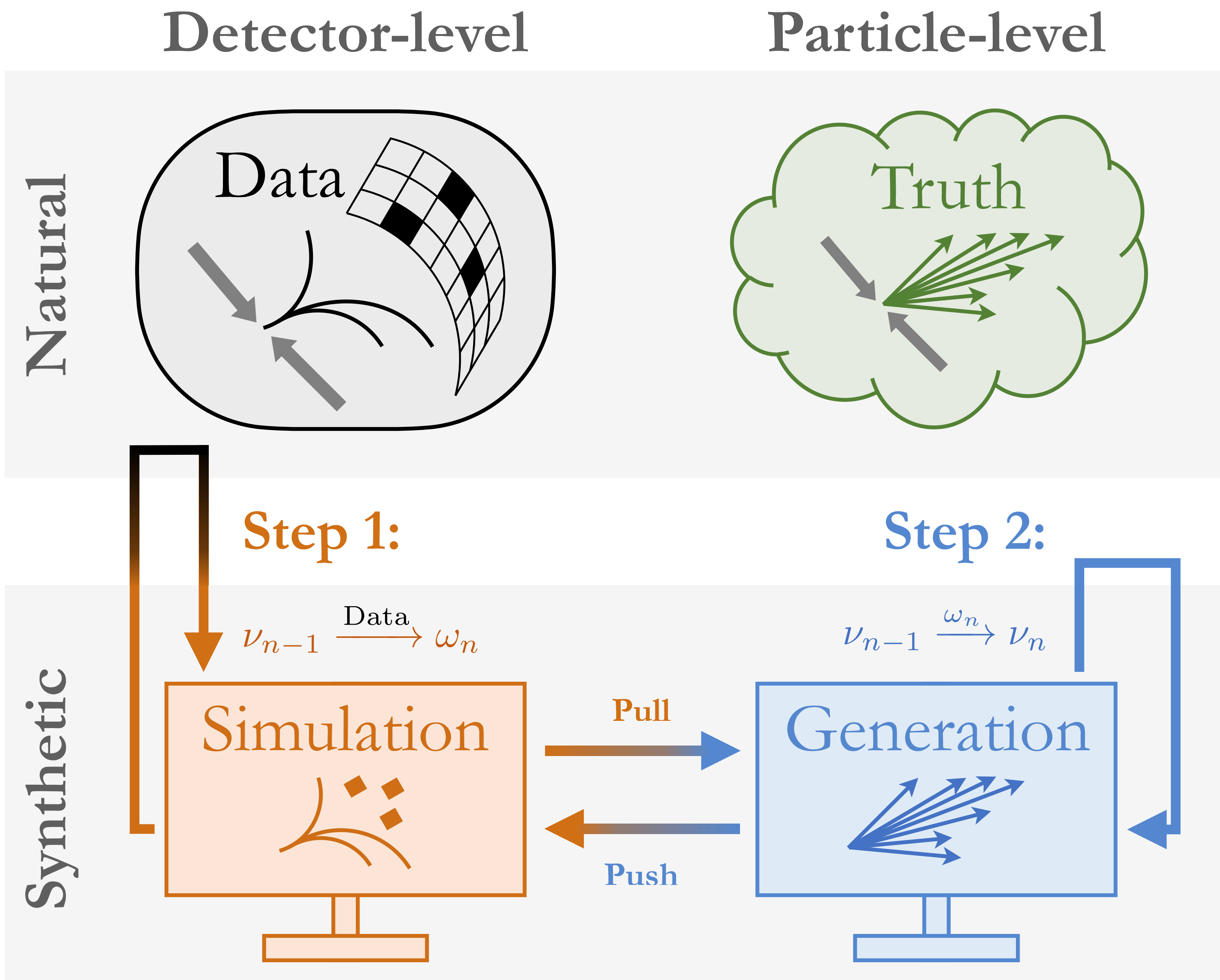


Step 1 – Reweights Sim_{n-1} to data, pulls weights back to particle-level Gen_{n-1}

OmniFold Algorithm – Schematic

[Andreassen, PTK, Metodiev, Nachman, Thaler, 1911.09107]

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}

Step 2 – Reweights Gen_{n-1} to (step 1)-weighted gen_{n-1}, pushes weights to detector-level Sim_n

OmniFold Algorithm – Equations

[Andreassen, PTK, Metodiev, Nachman, Thaler, 1911.09107]

Inputs

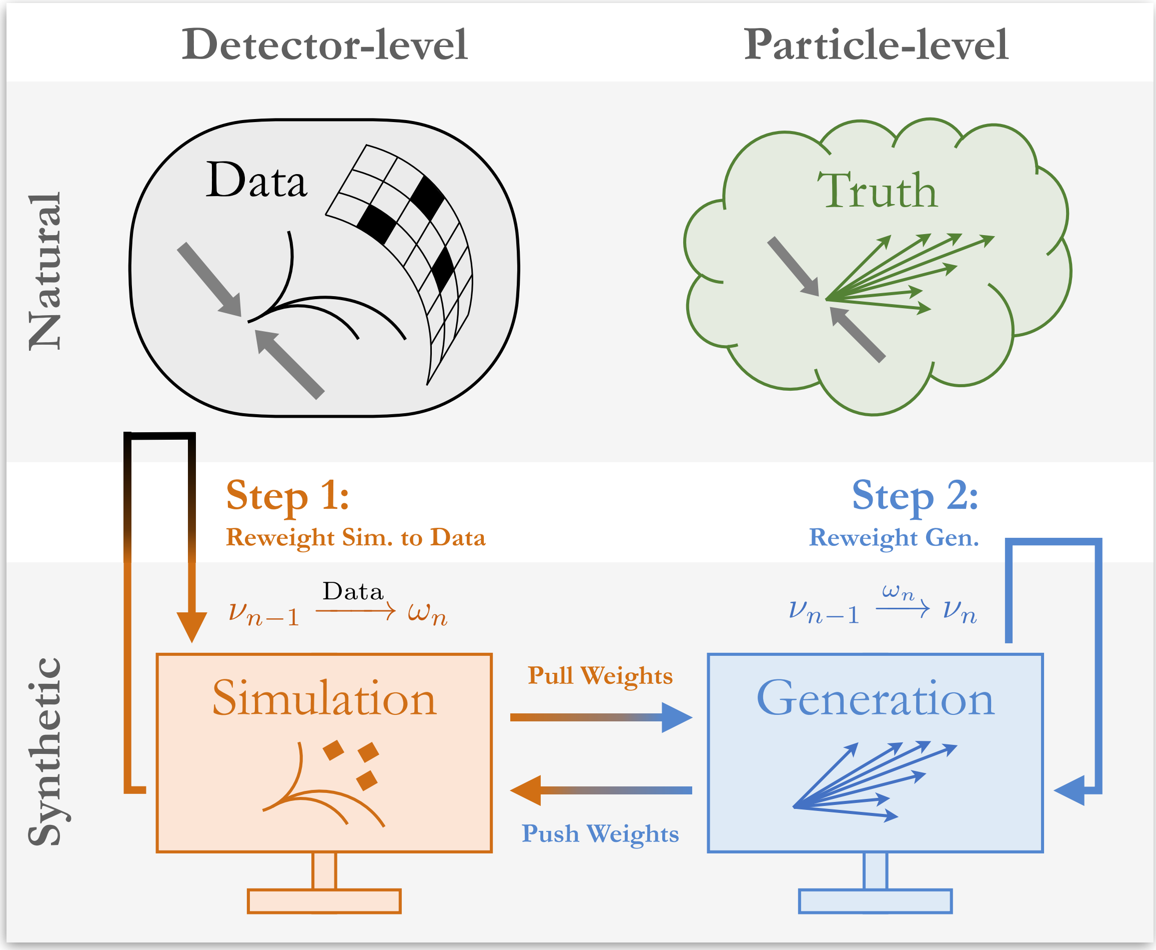
(t, m) – pairs of Gen and Sim events
$\nu_0(t)$ – initial particle-level weights for Gen
– Data

Results of Steps 1 and 2

$\nu_n(t)$ – particle-level weights for Gen, n^th iteration
$\omega_n(m)$ – detector-level weights for Sim, n^th iteration

Pulling/Pushing Weights

$\omega_n^{\text{pull}}(t) = \omega_n(m)$ – pulling ω_n back to particle-level
$\nu_n^{\text{push}}(m) = \nu_n(t)$ – pushing ν_n to detector-level



OmniFold Algorithm – Equations

[Andreassen, PTK, Metodiev, Nachman, Thaler, 1911.09107]

Inputs

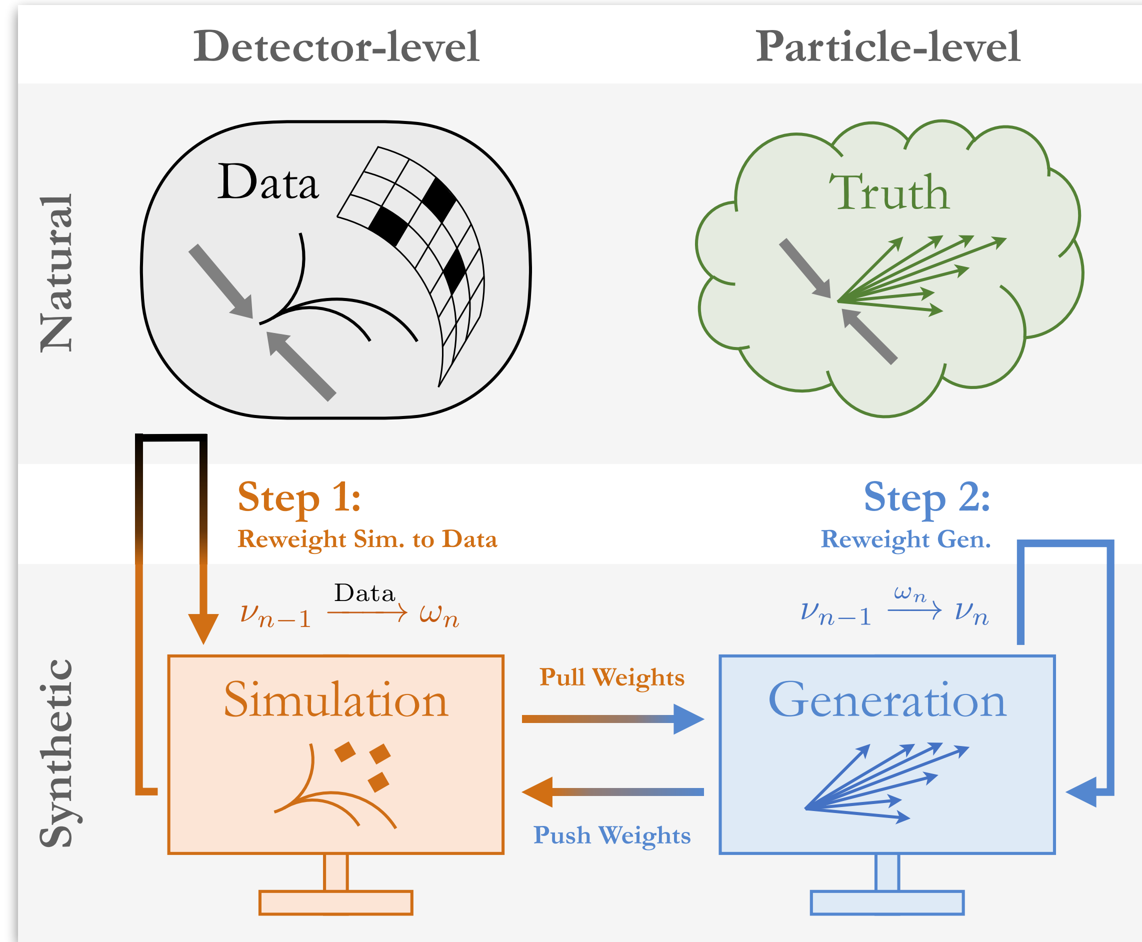
(t, m) – pairs of Gen and Sim events
$\nu_0(t)$ – initial particle-level weights for Gen
– Data

Results of Steps 1 and 2

$\nu_n(t)$ – particle-level weights for Gen, n[th] iteration
$\omega_n(m)$ – detector-level weights for Sim, n[th] iteration

Pulling/Pushing Weights

$\omega_n^{\text{pull}}(t) = \omega_n(m)$ – pulling ω_n back to particle-level
$\nu_n^{\text{push}}(m) = \nu_n(t)$ – pushing ν_n to detector-level



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

OmniFold Algorithm – Equations

[Andreassen, PTK, Metodiev, Nachman, Thaler, 1911.09107]

Inputs

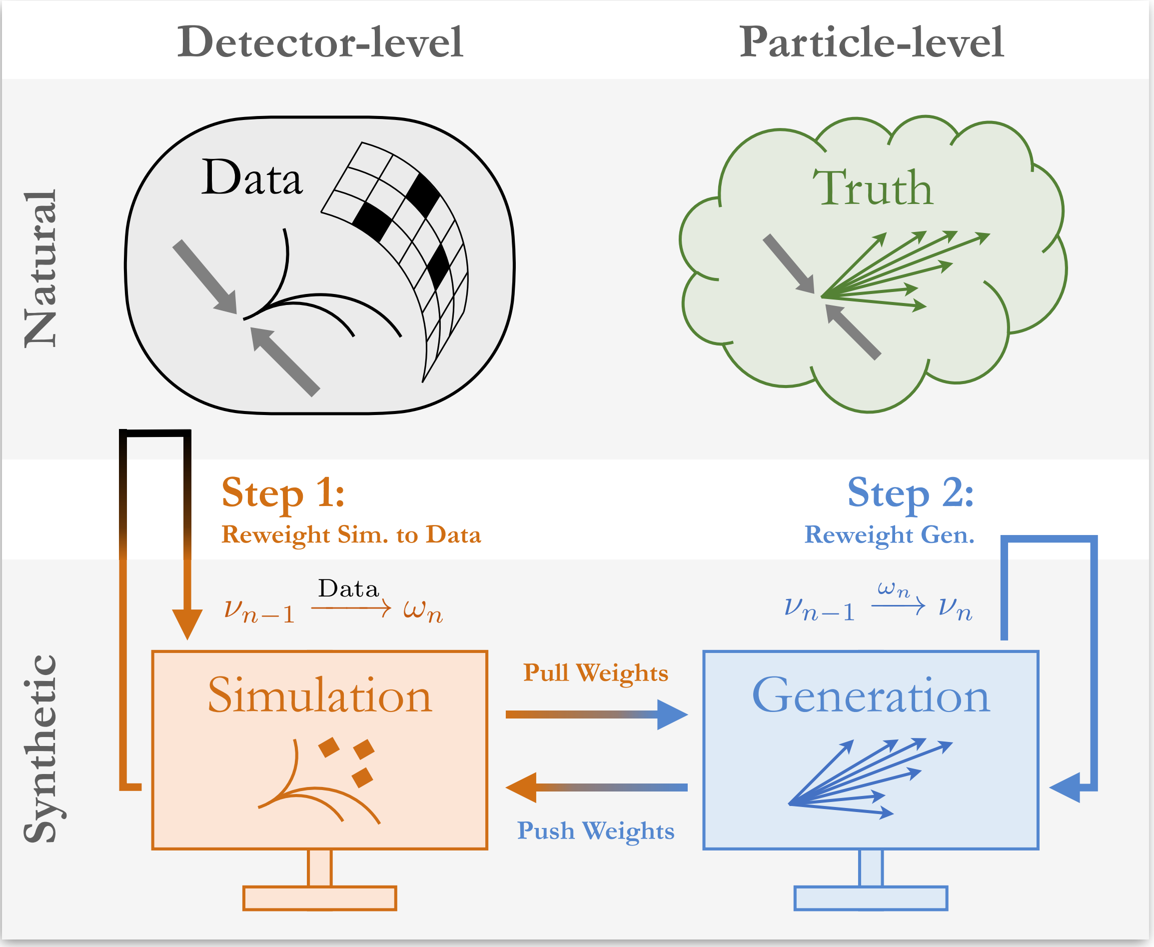
(t, m) – pairs of Gen and Sim events
$\nu_0(t)$ – initial particle-level weights for Gen
– Data

Results of Steps 1 and 2

$\nu_n(t)$ – particle-level weights for Gen, n$^{\text{th}}$ iteration
$\omega_n(m)$ – detector-level weights for Sim, n$^{\text{th}}$ iteration

Pulling/Pushing Weights

$\omega_n^{\text{pull}}(t) = \omega_n(m)$ – pulling ω_n back to particle-level
$\nu_n^{\text{push}}(m) = \nu_n(t)$ – pushing ν_n to detector-level



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 $\nu_n(t)$

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

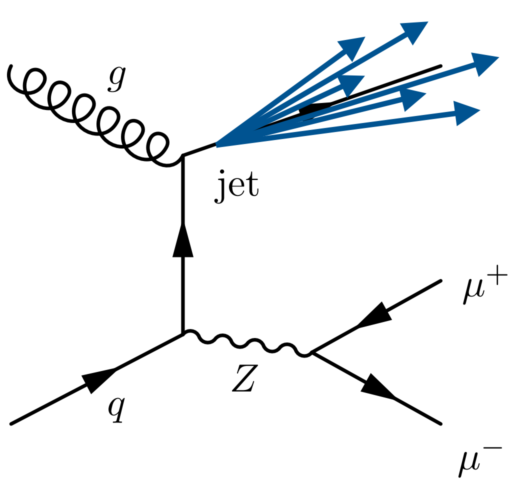
OmniFold is continuous IBU!

(See backup for IBU details)

After first iteration, with $\nu_0(t) = 1$:

$$\nu_1(t) p_{\text{Gen}}(t) = \int dm\, p_{\text{Gen}|\text{Sim}}(t|m)\, p_{\text{Data}}(m)$$

*Observables should be chosen responsibly

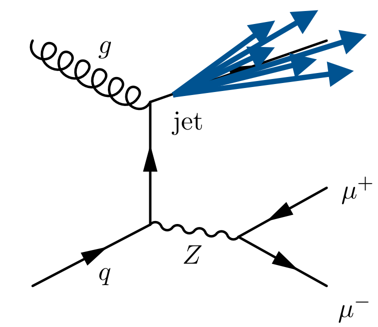


Unfolding Basics

OmniFold

Z + Jet Case Study

Ingredients for Z + Jet Case Study



$Z(\to \mu^+\mu^-)$ + 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

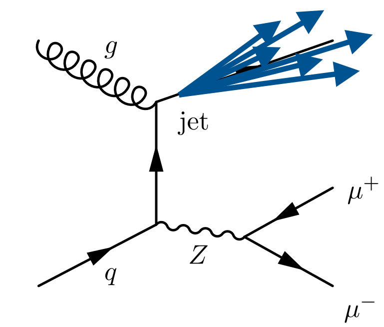
Jets

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

Ingredients for Z + Jet Case Study



$Z(\to \mu^+\mu^-)$ + Jet Events

"Data" – HERWIG 7.1.5
MC – PYTHIA 8.243, tune 26
1.6 million events each after cuts

Detector Simulation

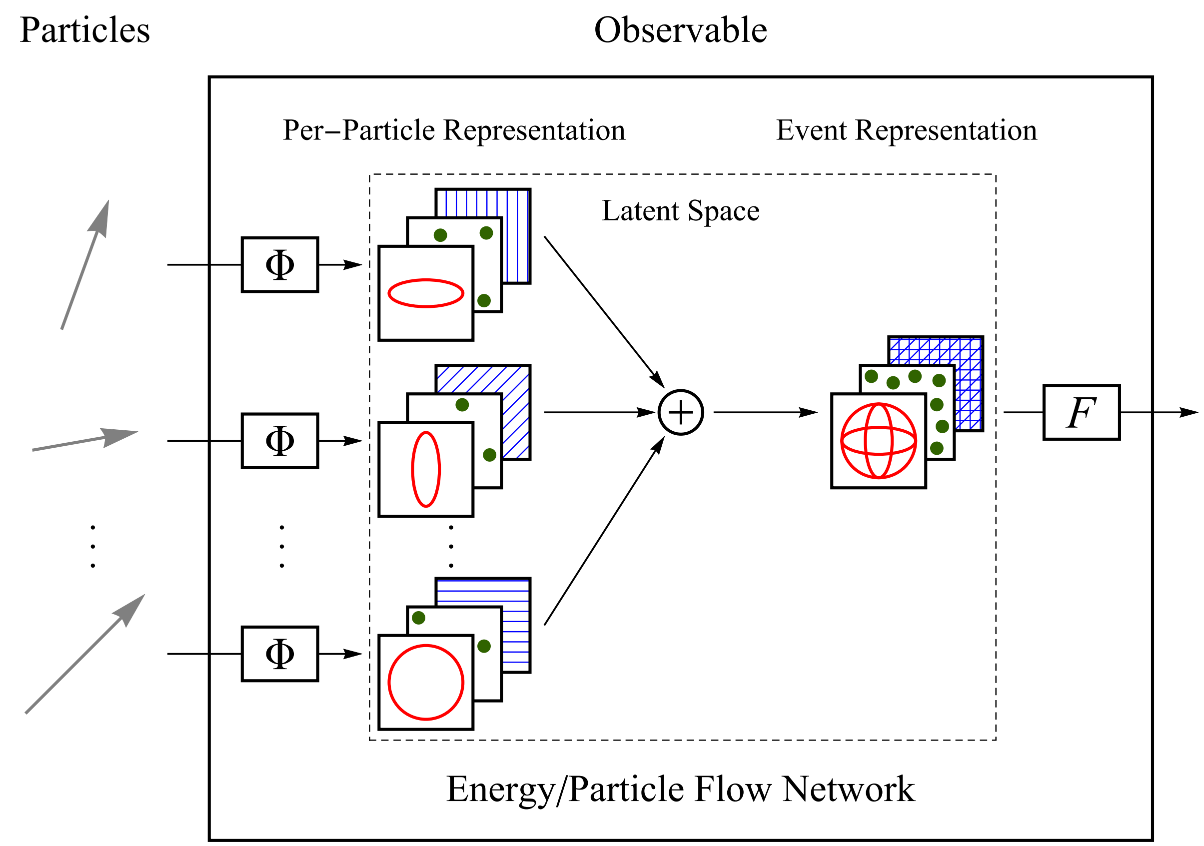
CMS-like detector – DELPHES 3.4.2

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

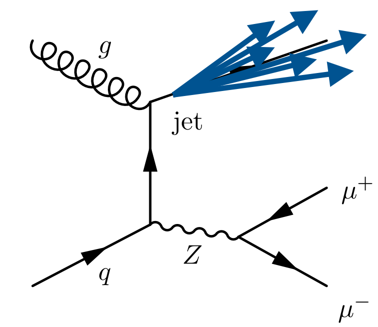


[PTK, Metodiev, Thaler, 1810.05165]
[PTK, Talk at ML4Jets 2018]

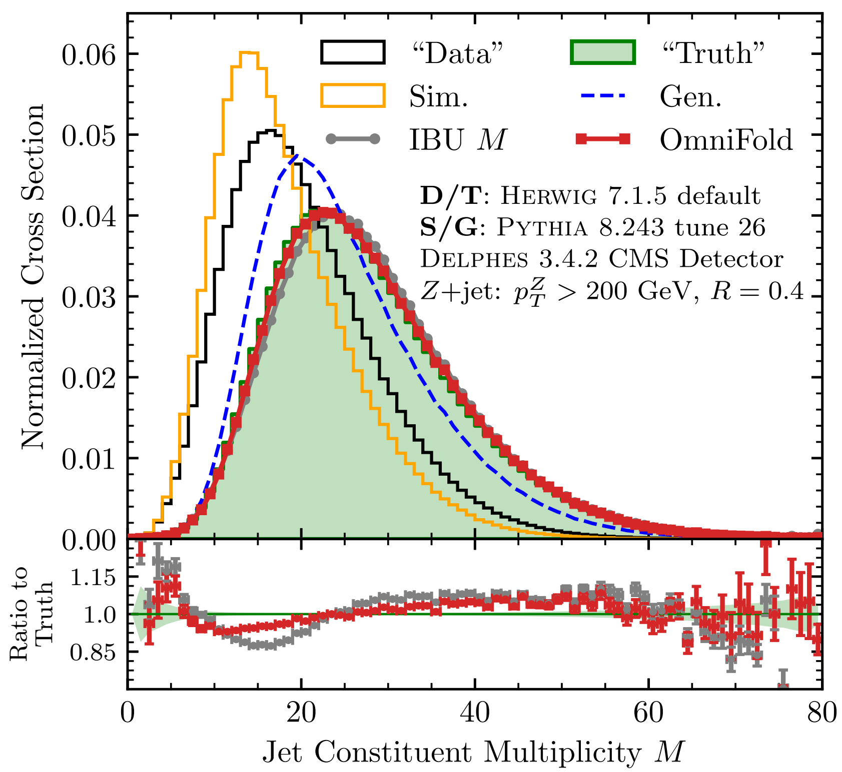
Particle Flow Network (PFN) architecture processes full radiation pattern of the event

- PFN-Ex: $(p_T, y, \phi,$ 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

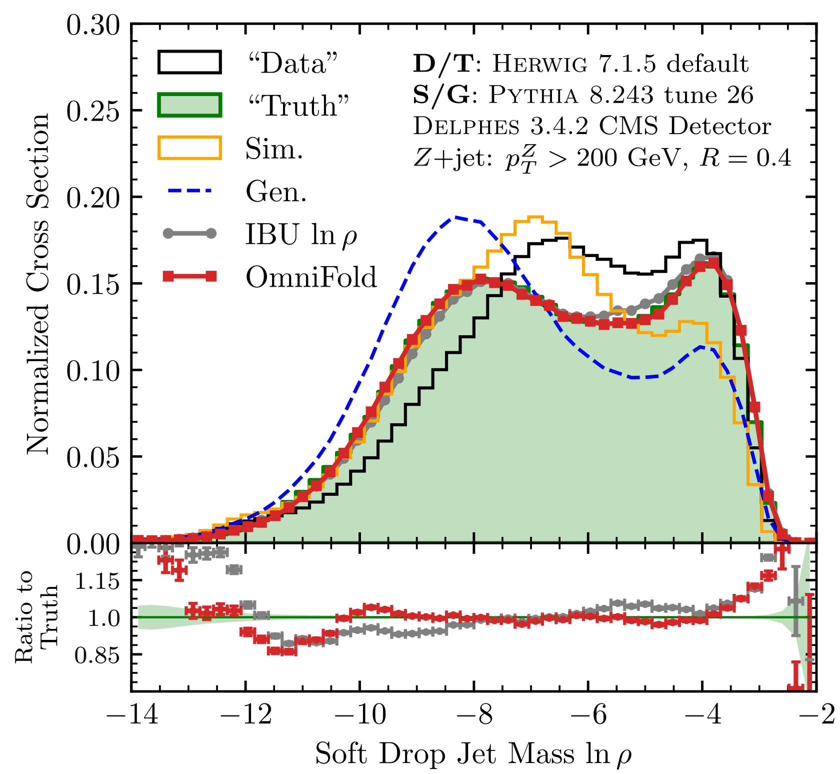
OmniFolding Jet Substructure Observables



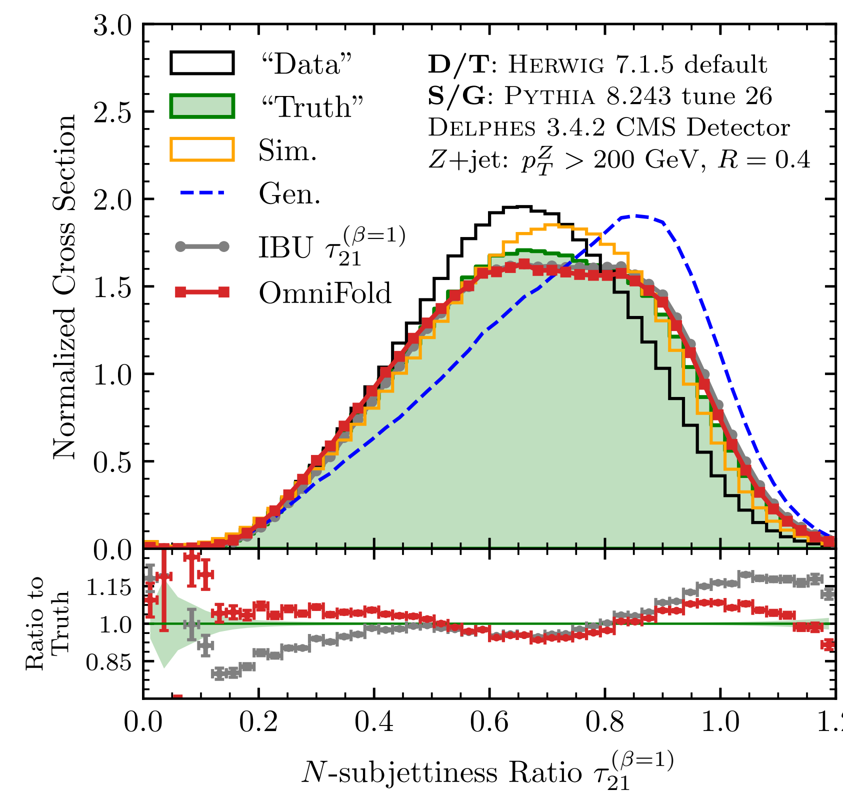
Single OmniFold instantiation vs. individual applications of IBU



IRC unsafe



IRC safe



Sudakov safe

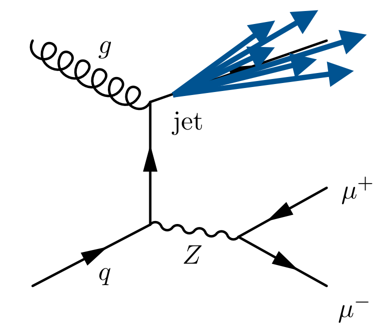
OmniFold equals or outperforms IBU

Five unfolding iterations in all cases

Statistical uncertainties on prior shown in ratio

(See backup for more distributions)

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

	Observable					
Method	m	M	w	$\ln \rho$	τ_{21}	z_g
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

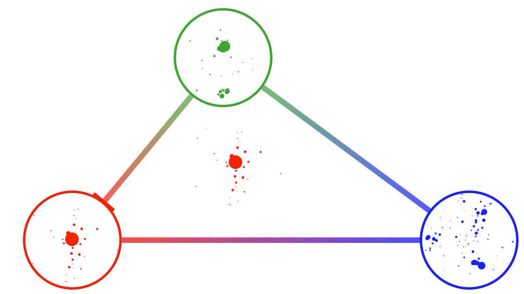
$$\Delta(p,q) = \frac{1}{2}\int d\lambda \, \frac{(p(\lambda)-q(\lambda))^2}{p(\lambda)+q(\lambda)} (\times 10^3)$$

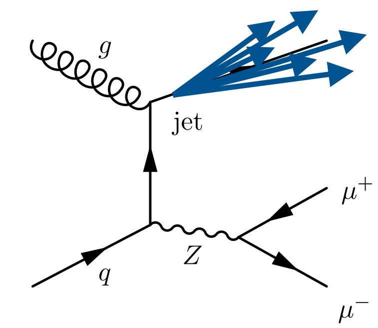
Single MULTIFOLD training based on all six observables

UNIFOLD is similar to or outperforms IBU

OMNIFOLD/MULTIFOLD outperform IBU on all observables!

Unfolding Beyond Observables

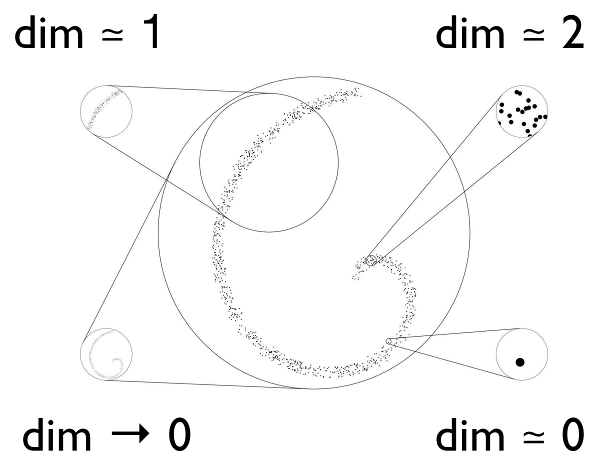




Energy Mover's Distance (EMD)
is a metric on the space of events

[PTK, Metodiev, Thaler, 1902.02346]

EMD enables calculating
correlation dimension of jets

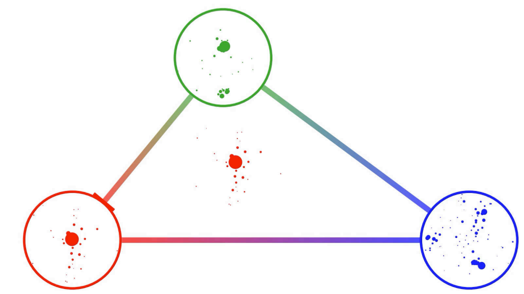


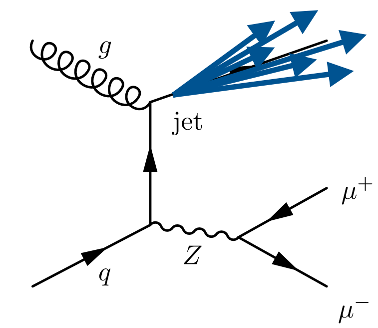
$$\dim(Q) = Q\frac{\partial}{\partial Q}\ln\sum_i\sum_j \boxed{w_i w'_j}\,\Theta(\mathrm{EMD}(\mathcal{E}_i,\mathcal{E}'_j) < Q)$$

Weighted events naturally accommodated

(See yesterday's talks by Eric Metodiev and Jack Collins)

Unfolding Beyond Observables



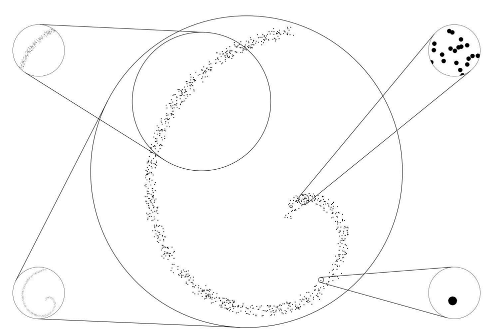


Energy Mover's Distance (EMD)
is a metric on the space of events

[PTK, Metodiev, Thaler, 1902.02346]

EMD enables calculating
correlation dimension of jets

dim ≃ 1 dim ≃ 2



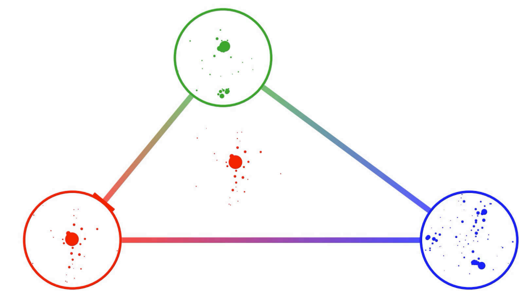
dim → 0 dim ≃ 0

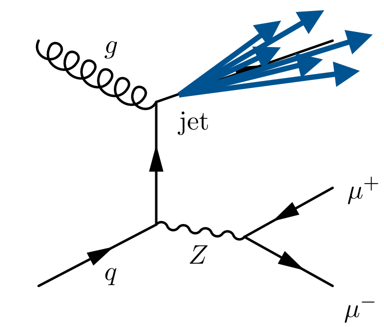
$$\dim(Q) = Q\frac{\partial}{\partial Q}\ln\sum_i\sum_j \boxed{w_i w'_j}\,\Theta(\mathrm{EMD}(\mathcal{E}_i,\mathcal{E}'_j) < Q)$$

Weighted events naturally accommodated

(See yesterday's talks by Eric Metodiev and Jack Collins)

Unfolding Beyond Observables



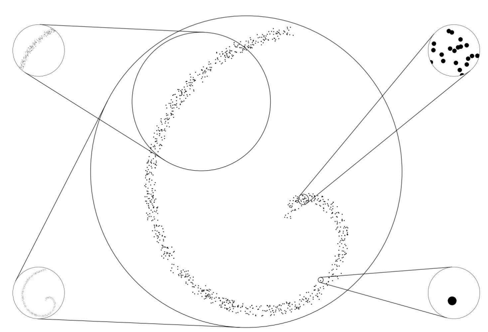


Energy Mover's Distance (EMD)
is a metric on the space of events

[PTK, Metodiev, Thaler, 1902.02346]

EMD enables calculating
correlation dimension of jets

dim ≃ 1 dim ≃ 2

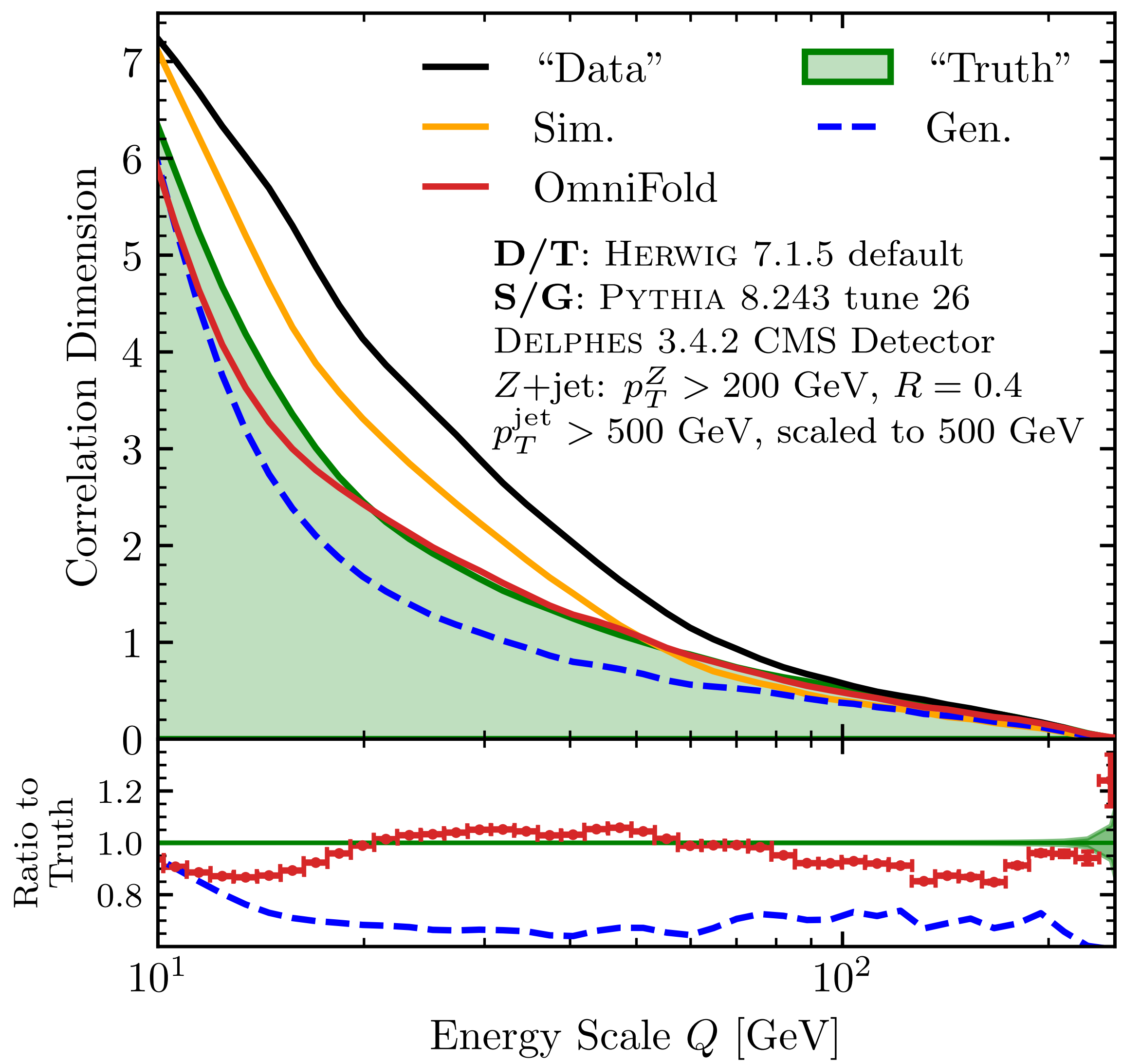


dim → 0 dim ≃ 0

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

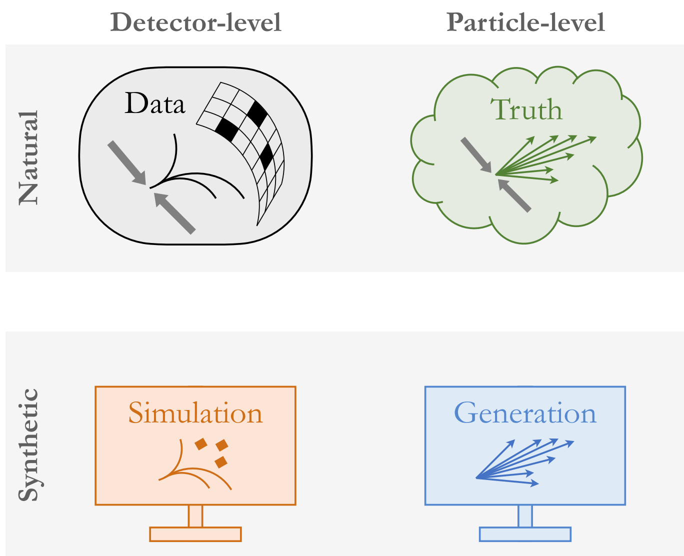
Weighted events naturally accommodated

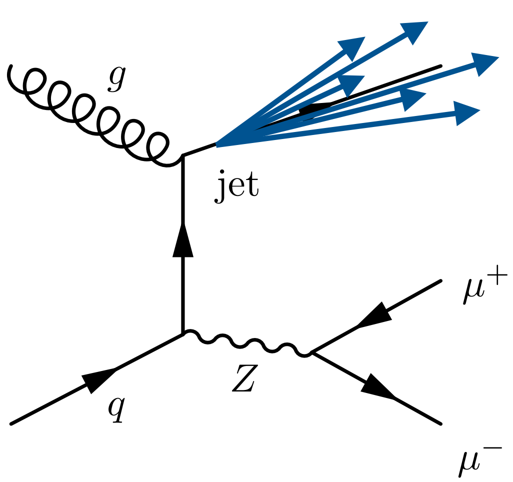
(See yesterday's talks by Eric Metodiev and Jack Collins)



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

Larger detector effects and loss of stats seen at low Q





Unfolding Basics

Measurements are unfolded to mitigate detector effects

Standard unfolding is binned and low-dimensional

OmniFold

ML-based method simultaneously unfolds all observables

Unbinned, full phase-space information

Z + Jet Case Study

MC study of a realistic measurement with public datasets

OmniFold, MultiFold, UniFold ready for action

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

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

Kluge **OmniFold** 3000 Automatic Folding and Gluing System

Additional Slides

Future Directions

Challenges

Dealing with detector inefficiencies

Apply more restrictive cuts after unfolding

Include gen/sim pairs with empty events in the training

Systematic uncertainties

Existing strategies should reasonably carry over

Parametrize high-dimensional systematic uncertainties, see [Nachman, 1909.03081]

Opportunities

Training ML models on unfolded data

OmniFold allows any model that can be defined on weighted data to be unfolded

OmniFold and CMS Open Data

CMS 2011A Jet dataset processed into simple to use HDF5 files

Iterated Bayesian Unfolding (IBU)

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

$R_{ij} = \Pr(\text{measure } i \mid \text{truth is } j)$ R is in general non-square and non-invertible

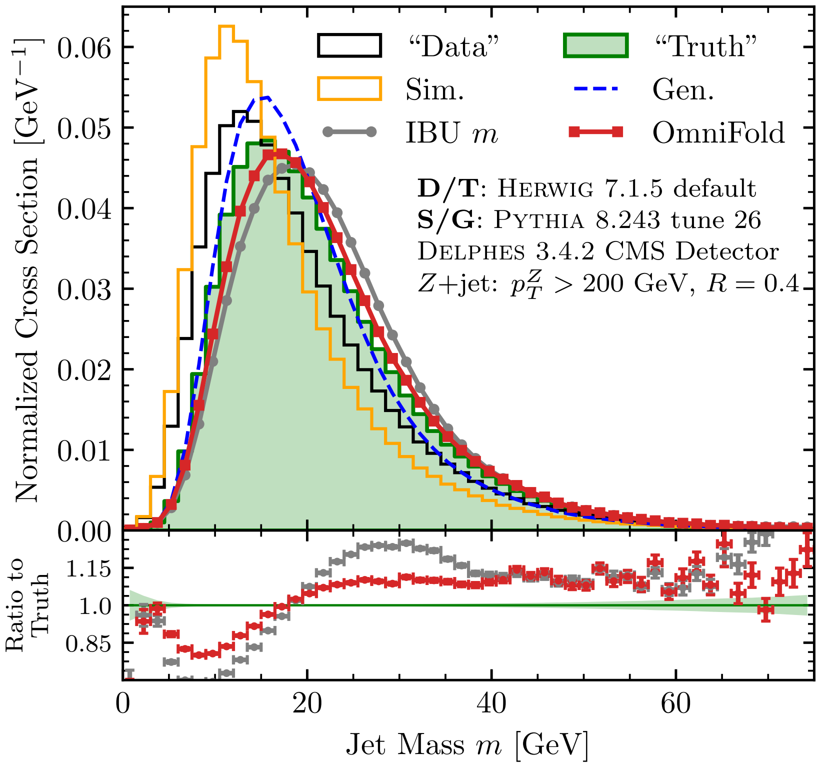
Calculate new particle-level distribution using Bayes' theorem

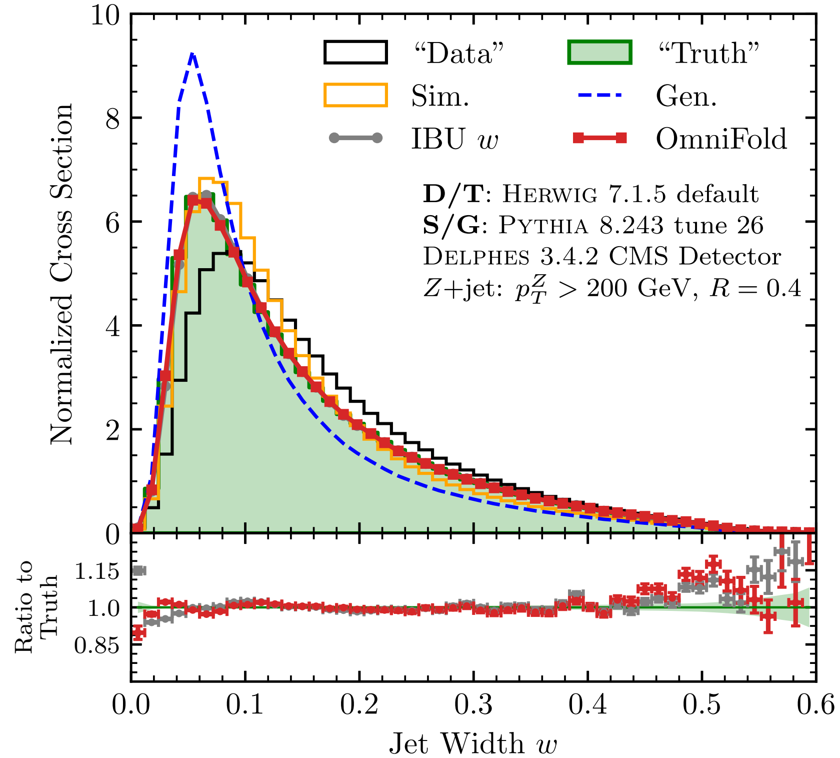
$$t_j^{(n)} = \sum_i \Pr(\text{truth}_{n-1} \text{ is } j \mid \text{measure } i) \times \Pr(\text{measure } i) = \sum_i \frac{R_{ij} t_j^{(n-1)}}{\sum_k R_{ik} t_k^{(n-1)}} \times m_i$$

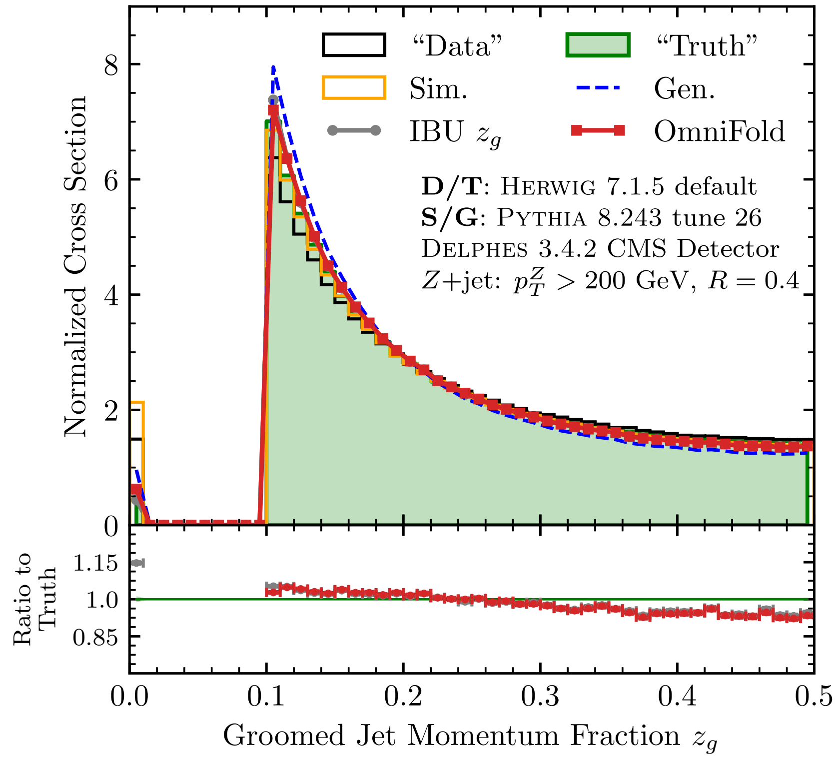
Iterate procedure to remove dependence on prior

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

Additional OmniFolded Distributions







Jet mass affected by particle masses

IRC-safe observables easier to unfold

z_g remarkably stable under choice of method

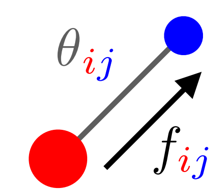
The Energy Mover's Distance (EMD)

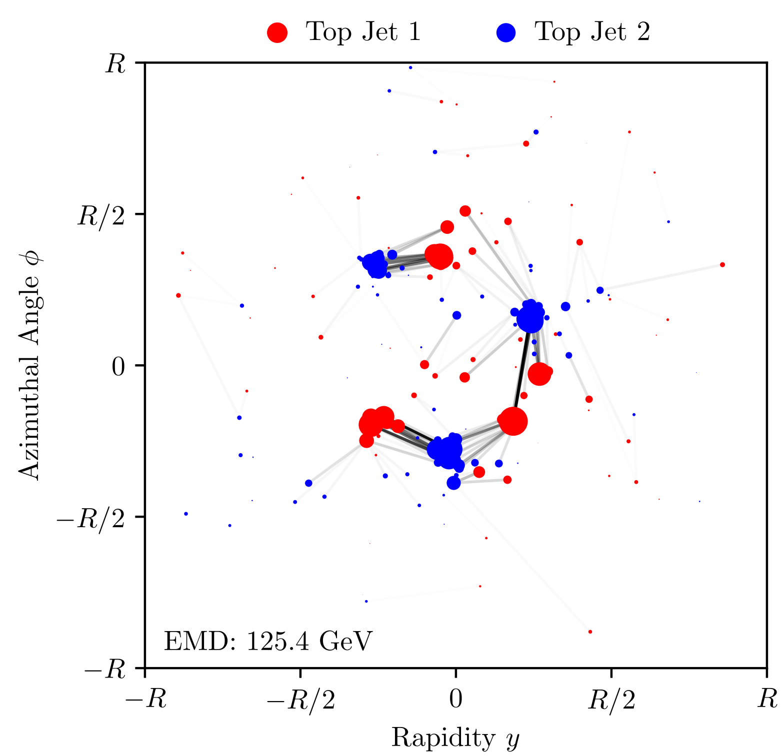
[PTK, Metodiev, Thaler, 1902.02346]

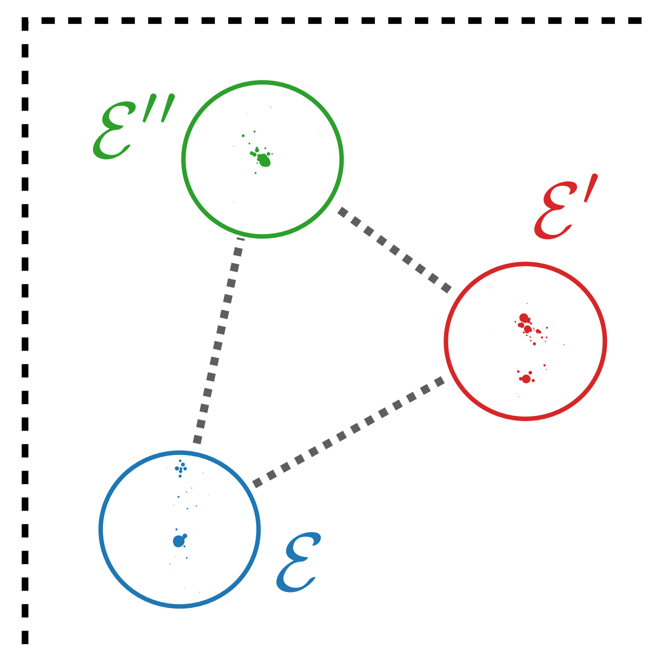
EMD between energy flows defines a metric on the space of events

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

$$\sum_j f_{ij} \leq E_i, \quad \sum_i f_{ij} \leq E'_j, \quad \sum_{ij} f_{ij} = \min\left(\sum_i E_i, \sum_j E'_j\right)$$





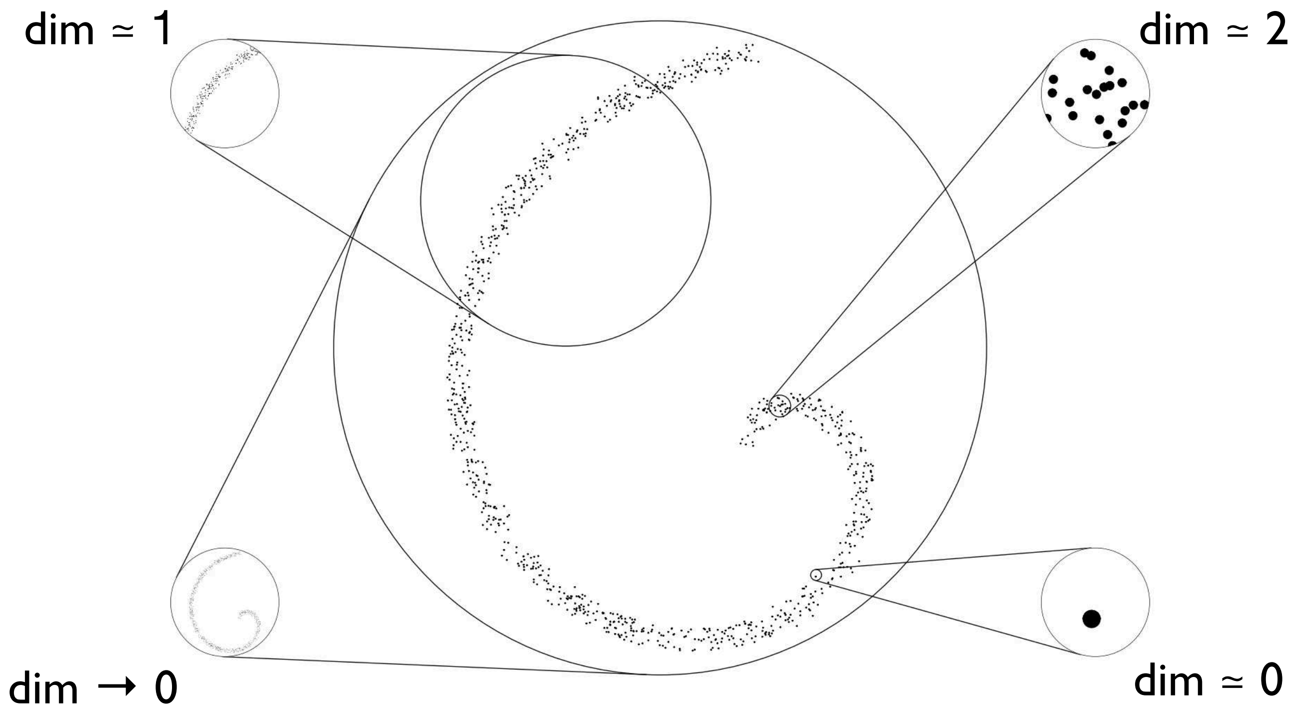


Triangle inequality satisfied for $R \geq d_{\max}/2$

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

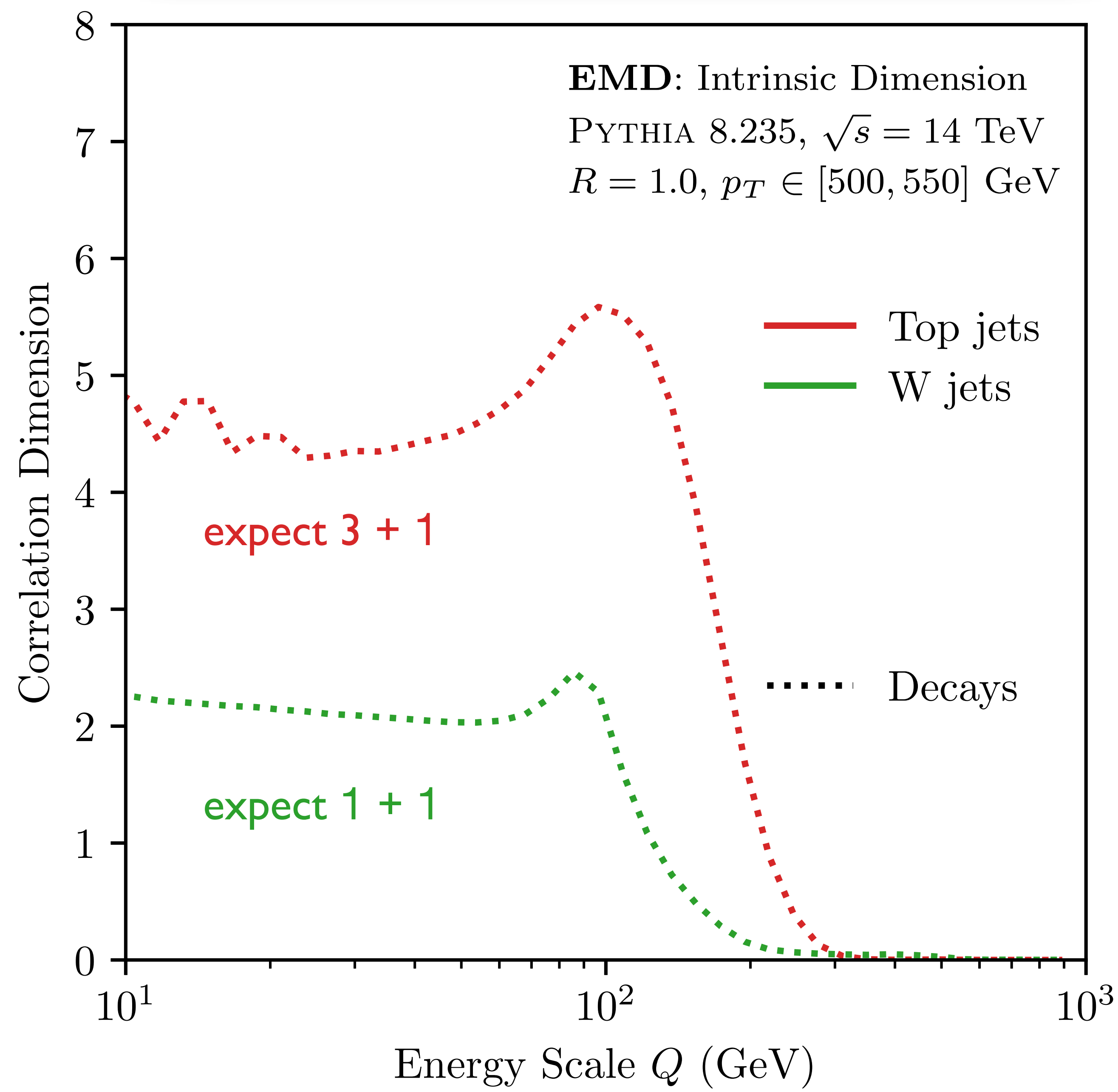
Manifold Dimensions of Event Space

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



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

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



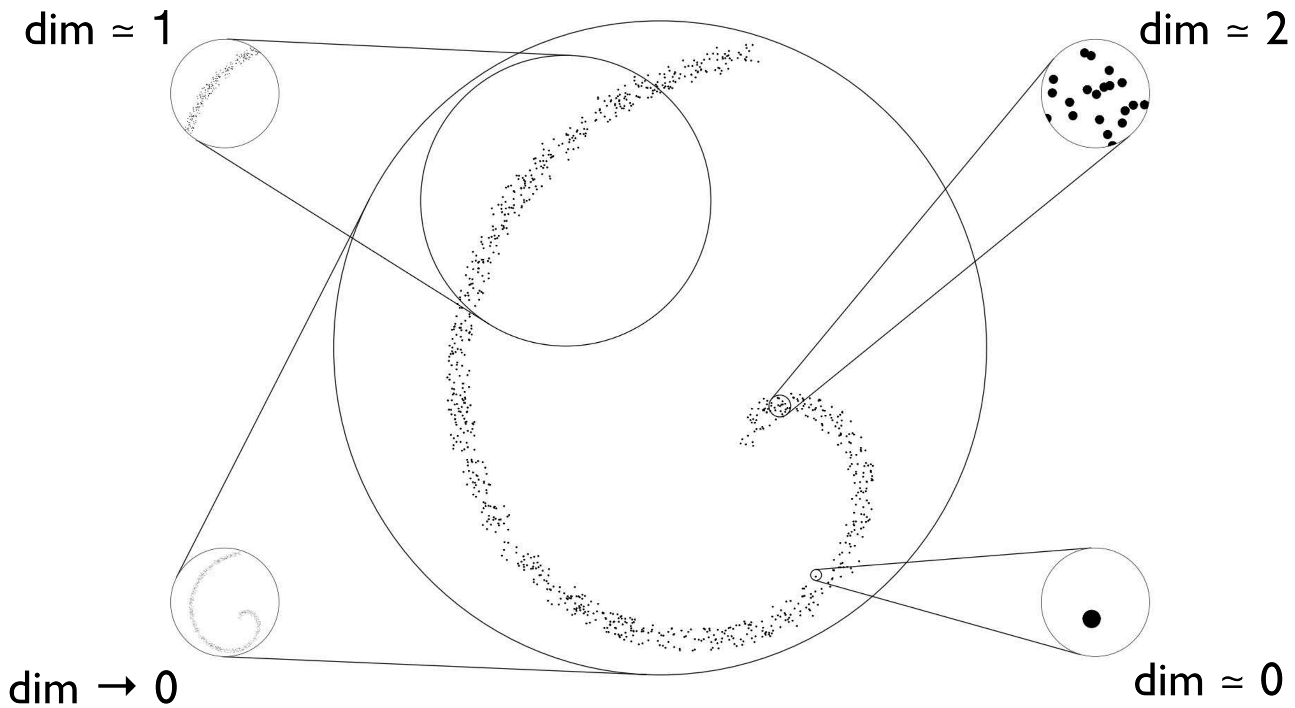
Correlation dimension lessons:

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

[Grassberger, Procaccia, PRL 1983; PTK, Metodiev, Thaler, 1902.02346]

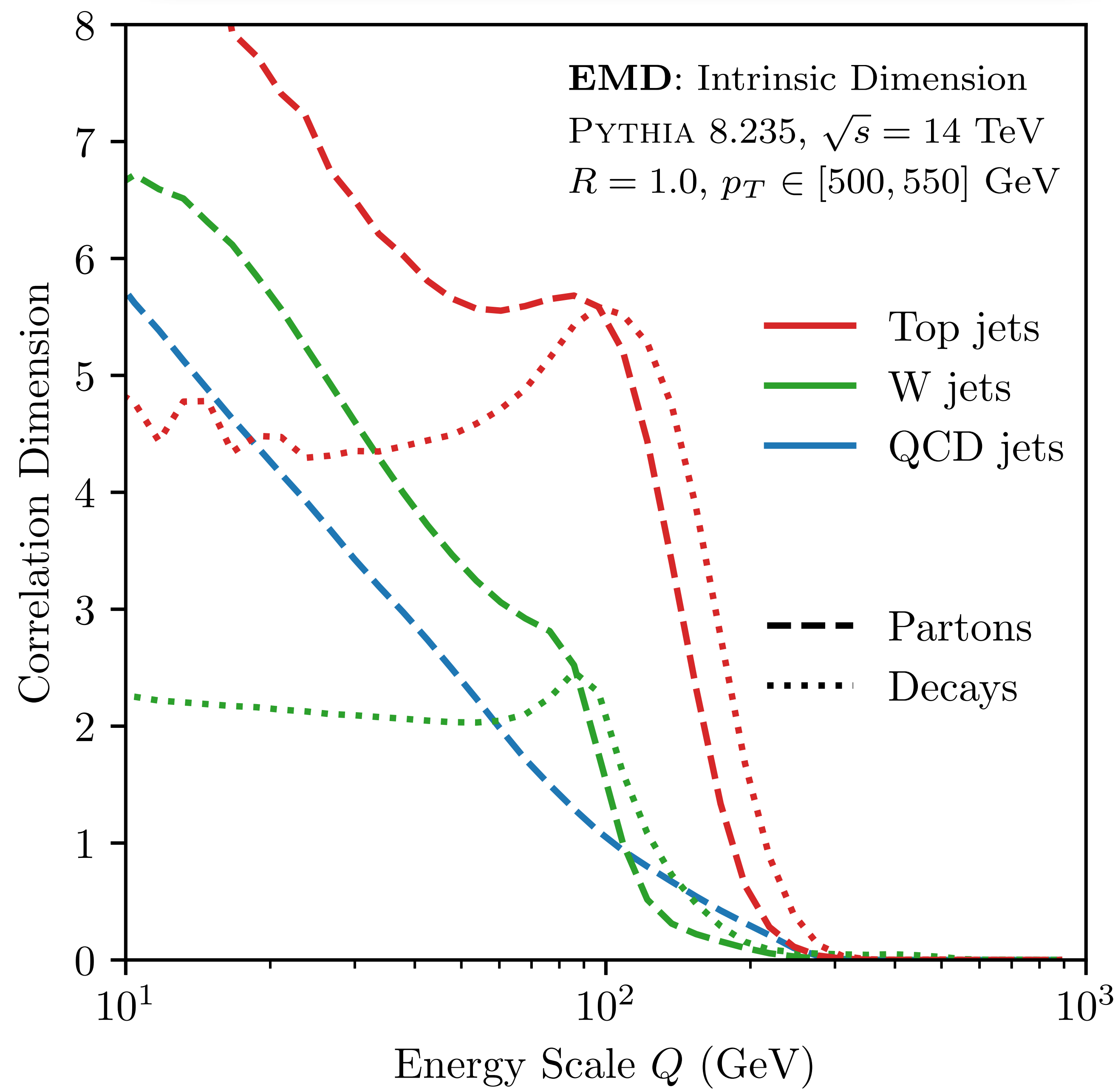
Manifold Dimensions of Event Space

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



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

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



Correlation dimension lessons:

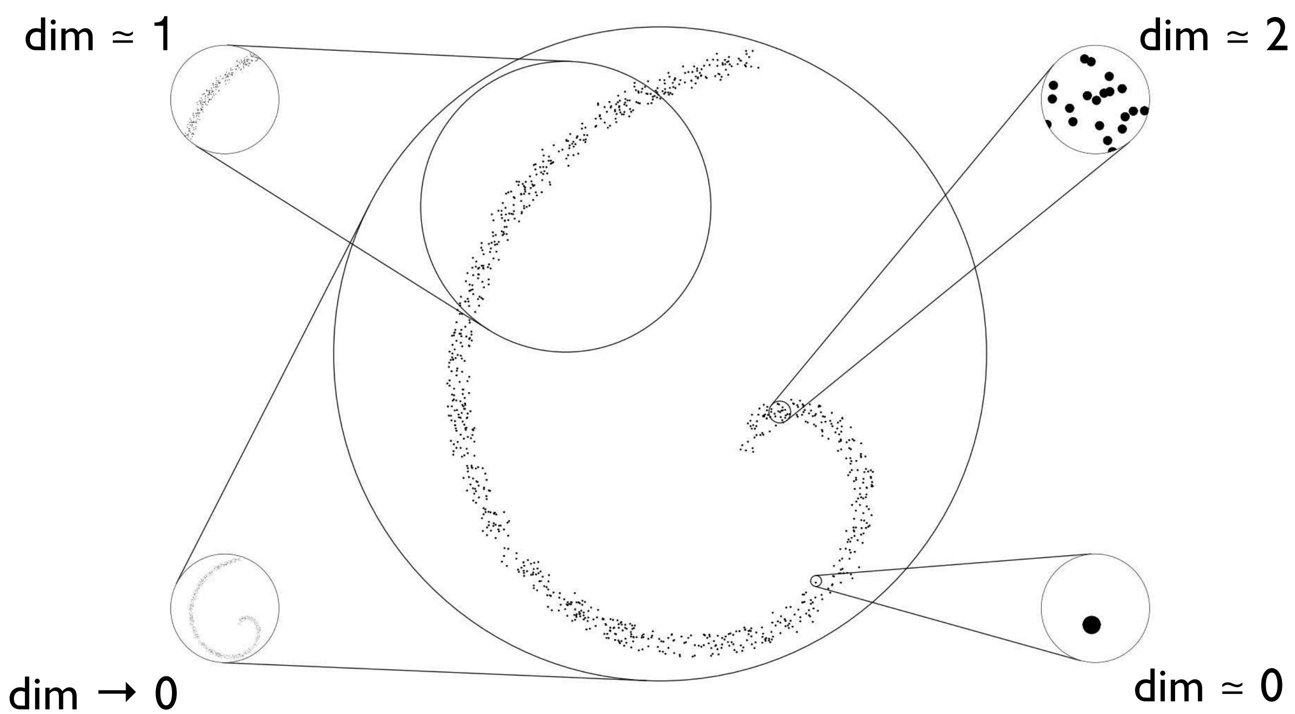
- Decays are "constant" dim. at low *Q*
- Complexity hierarchy: QCD < W < Top
- Fragmentation increases dim. at smaller scales

[Grassberger, Procaccia, PRL 1983; PTK, Metodiev, Thaler, 1902.02346]

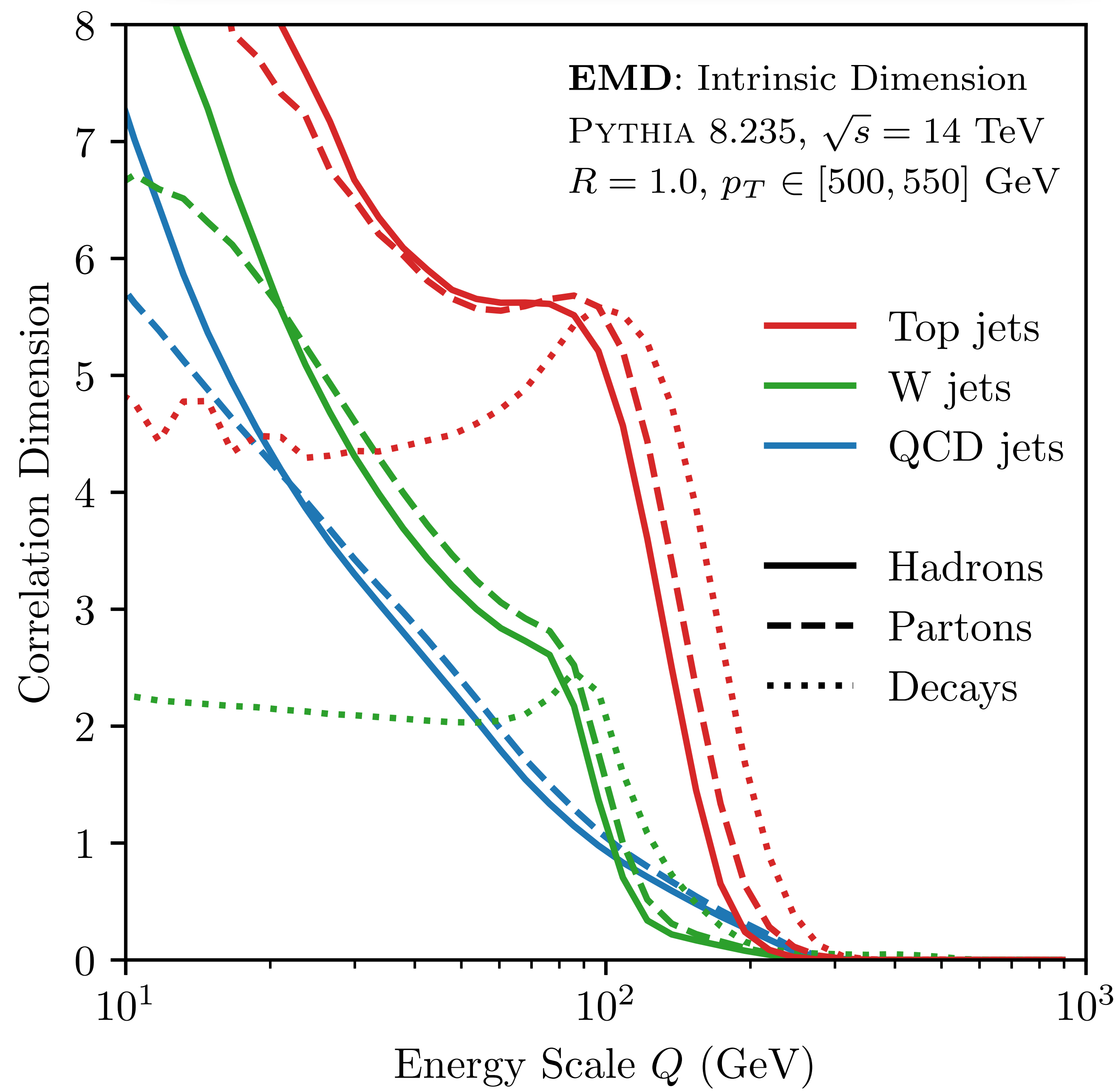
Manifold Dimensions of Event Space

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

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



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



Correlation dimension lessons:

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

[Grassberger, Procaccia, PRL 1983; PTK, Metodiev, Thaler, 1902.02346]

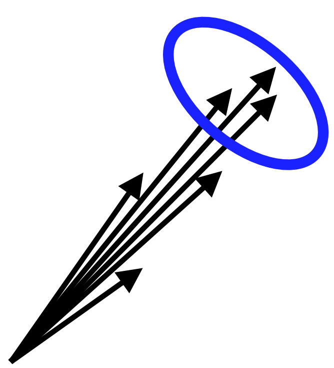
Quark and Gluon Correlation Dimensions

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

Leading log (single emission) calculation:

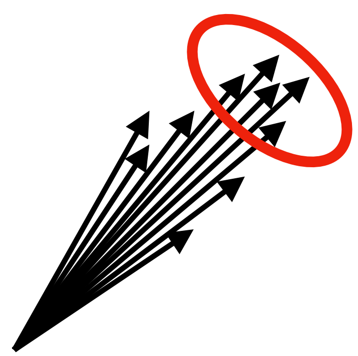
$$\dim_i(Q) \simeq -\frac{8\alpha_s}{\pi}C_i\ln\frac{Q}{p_T/2}$$

color factor



Quark

$C_F = 4/3$



Gluon

$C_A = 3$