Towards a data-driven model of hadronization

IAIFI JC

Tony Menzo

PhD candidate, University of Cincinnati

MLHAD

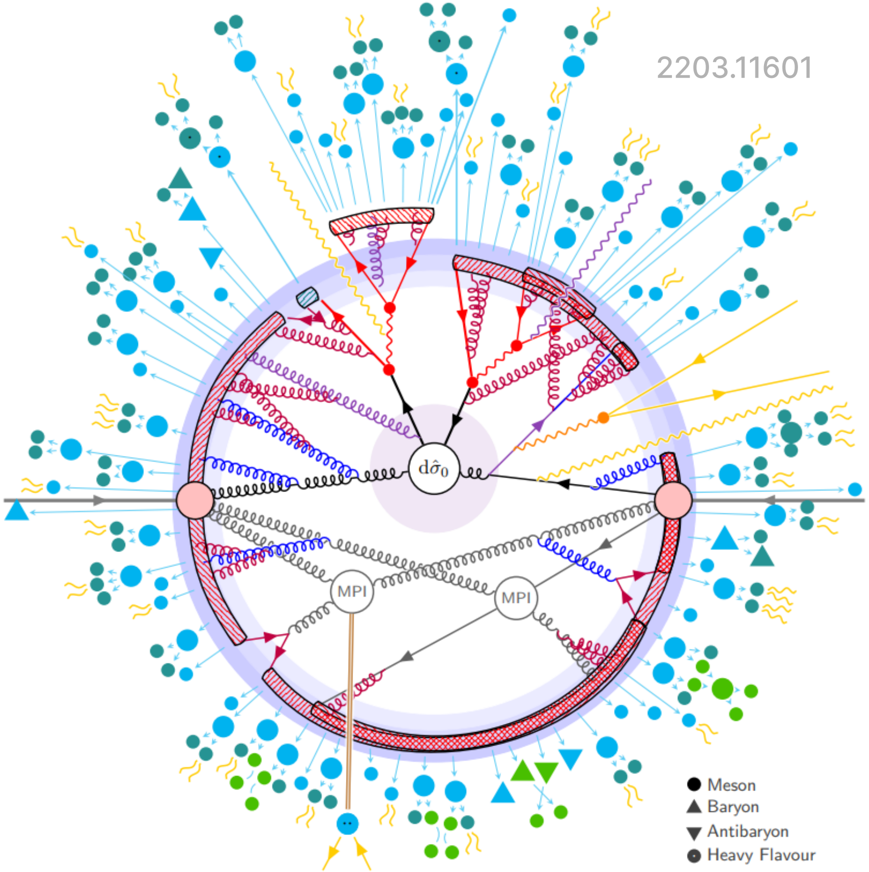
In collaboration with:

Christian Bierlich, Phil Ilten, Stephen Mrenna, Manuel Szewc, Michael Wilkinson, Ahmed Youssef, and Jure Zupan

Based upon work in 2203.04983, 2309.xxxxx

Hadronization

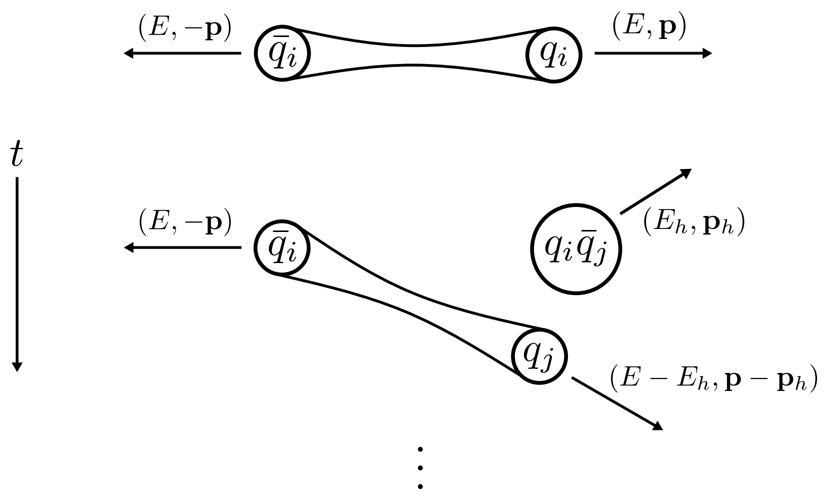
The conversion of a partonic system consisting of quarks, anti-quarks, and gluons ($q\bar{q}, qg\bar{q}, gg, qqq, qq\bar{q}\bar{q}, qqqgg, \ldots$) to a final state consisting of hadrons ($h_1, h_2, \ldots, h_n$).



Phenomenological models

Lund string model

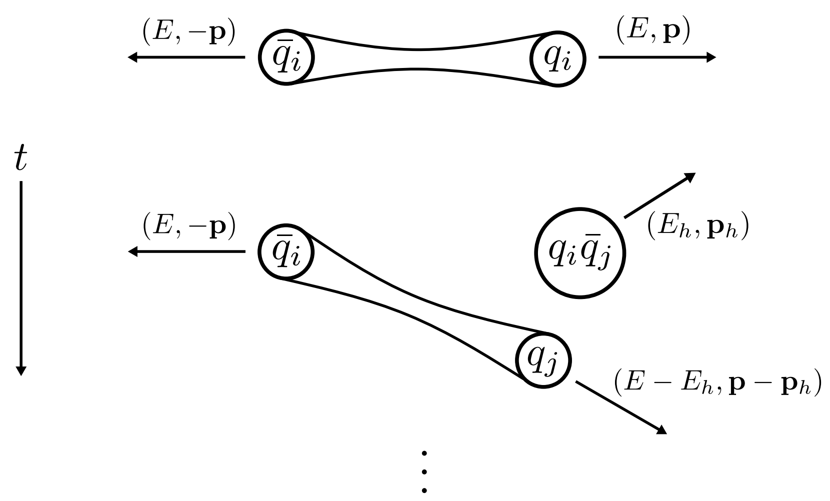
(used in Pythia)



Phenomenological models

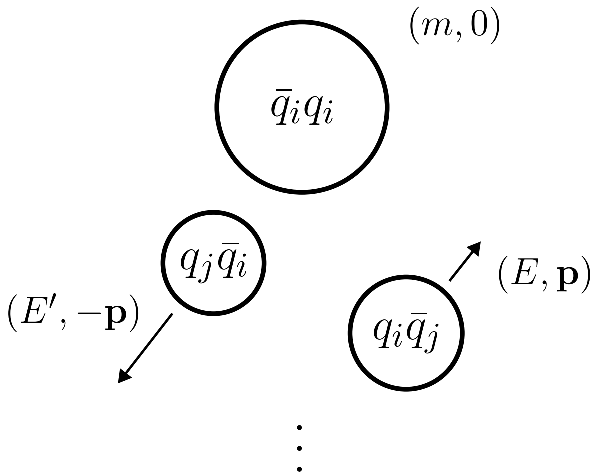
Lund string model

(used in Pythia)



Cluster model

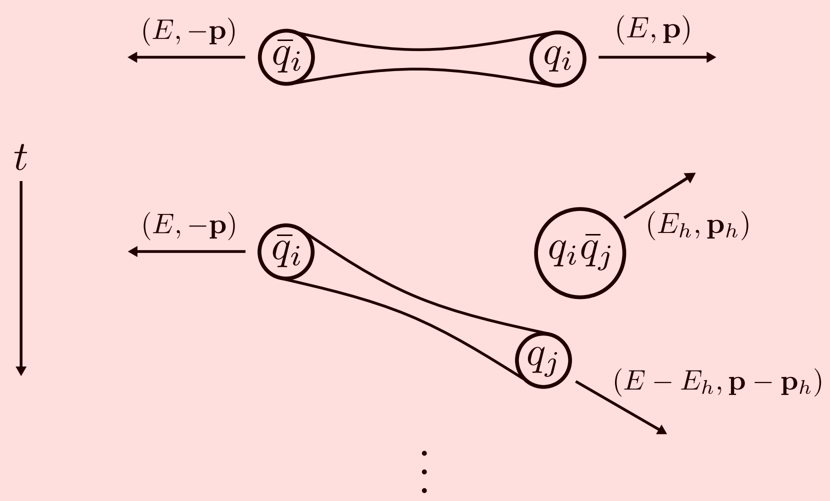
(used in Herwig)



Phenomenological models

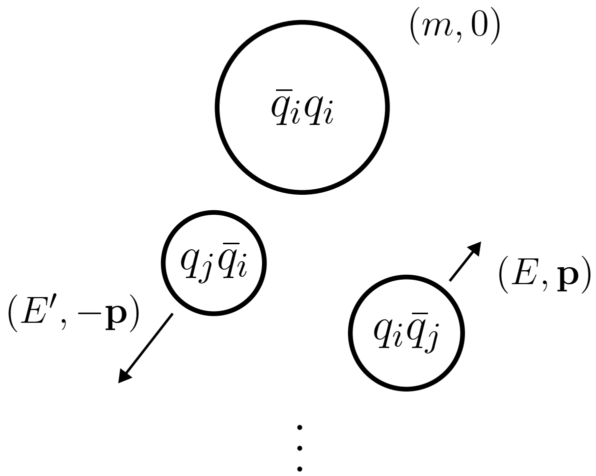
Lund string model

(used in Pythia)



Cluster model

(used in Herwig)



Stringy hadronization: overview

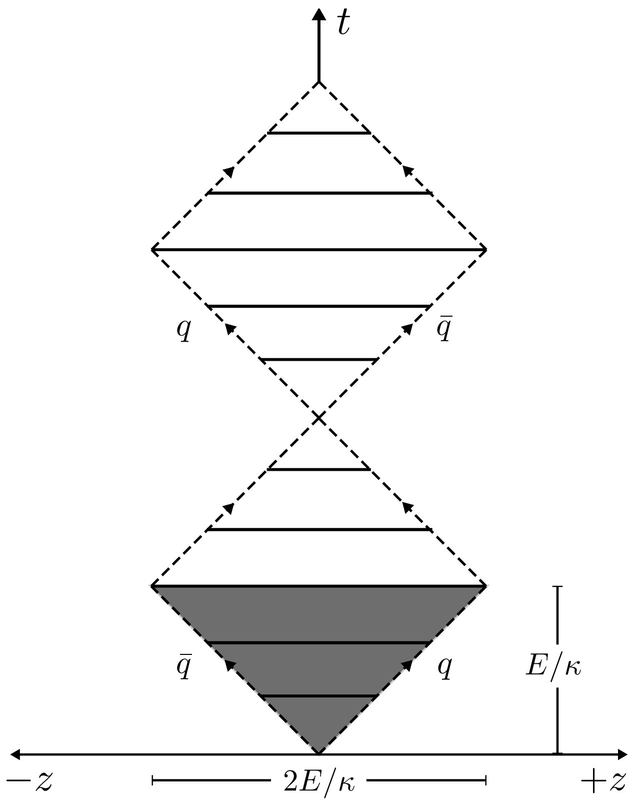
Consider the simplest hadronizing system:

A $q\bar{q}$ pair oriented along the z-axis, with equal and opposite momentum.

Treat this as a semi-classical system with potential

$$V(r) = \kappa r$$

A color flux tube forms between the $q\bar{q}$ pair and in the absence of string breaks the system follows a 'yo-yo' motion

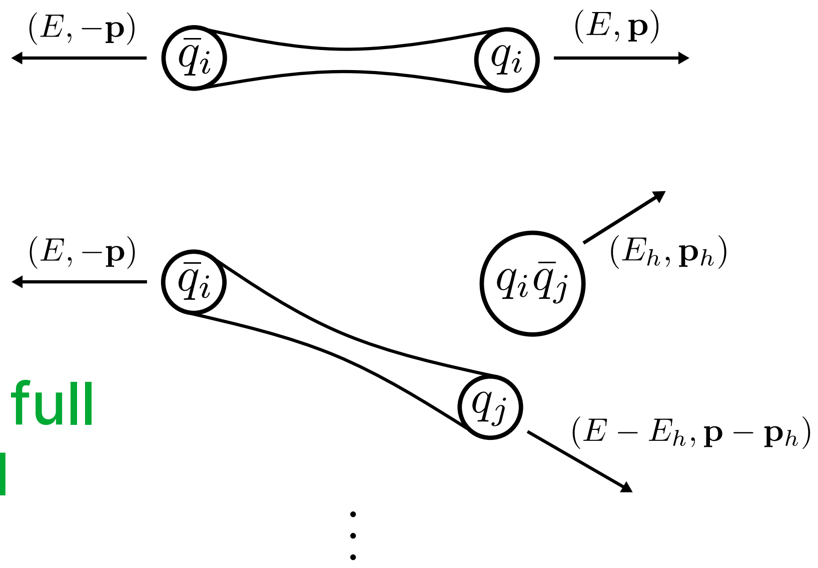


Stringy hadronization: overview

Consider the simplest hadronizing system:

A $q\bar{q}$ pair oriented along the z-axis, with equal and opposite momentum.

With string breaks the system will produce $q'\bar{q}'$ pairs out of the vacuum along the string.

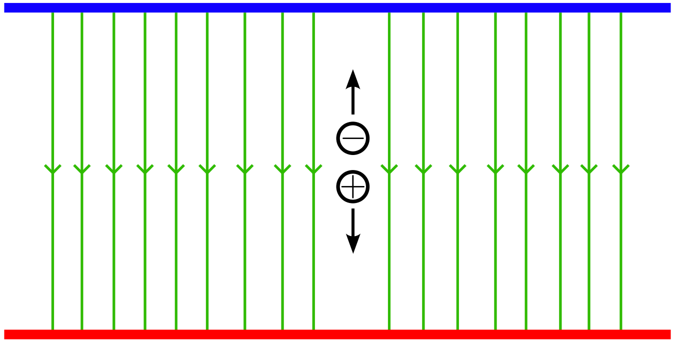


Causally disconnected fragmentations: the full hadronization event can be understood and implemented from the dynamics of a single emission

Stringy hadronization: kinematics

p_T

Transverse momentum of hadrons stems solely from the tunneling of $q'\bar{q}'$ pairs out of the vacuum breaking the string



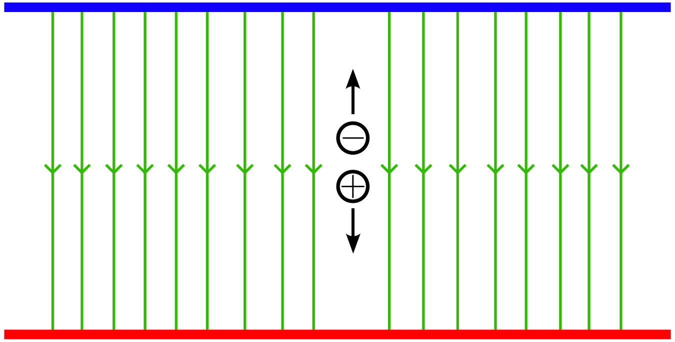
Implemented in practice by giving quarks gaussian p_T kicks

$$\mathcal{P}(p_x, p_y; \sigma_{p_T}) = \frac{1}{\pi \sigma_{p_T}^2} \exp\left(-\frac{p_x^2 + p_y^2}{\sigma_{p_T}^2}\right)$$

Stringy hadronization: kinematics

p_T

Transverse momentum of hadrons stems solely from the tunneling of $q'\bar{q}'$ pairs out of the vacuum breaking the string



Implemented in practice by giving quarks gaussian p_T kicks

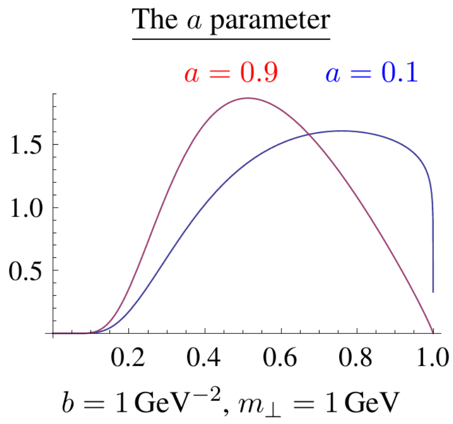
$$\mathcal{P}(p_x, p_y; \sigma_{p_T}) = \frac{1}{\pi\sigma_{p_T}^2} \exp\left(-\frac{p_x^2+p_y^2}{\sigma_{p_T}^2}\right)$$

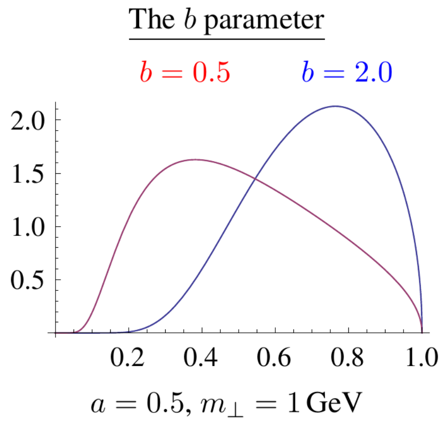
p_z

The momentum fraction z of each fragmenting hadron is sampled according to the

Lund fragmentation function

$$f(z) \propto \frac{(1-z)^a}{z} \exp\left(\frac{-bm_\perp^2}{z}\right), \qquad z = \frac{p_z + E_h}{2E}$$

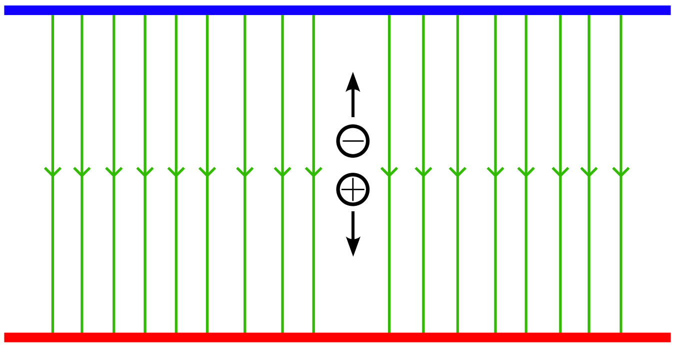




Stringy hadronization: kinematics

p_T

Transverse momentum of hadrons stems solely from the tunneling of q'$\overline{\text{q}}$' pairs out of the vacuum breaking the string



Implemented in practice by giving quarks gaussian p_T kicks

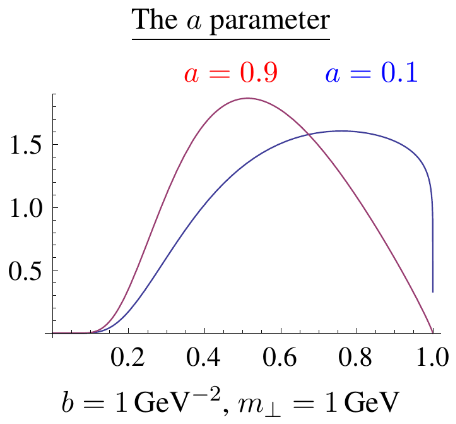
$$\mathcal{P}(p_x, p_y; \sigma_{p_T}) = \frac{1}{\pi \sigma_{p_T}^2} \exp\left(-\frac{p_x^2 + p_y^2}{\sigma_{p_T}^2}\right)$$

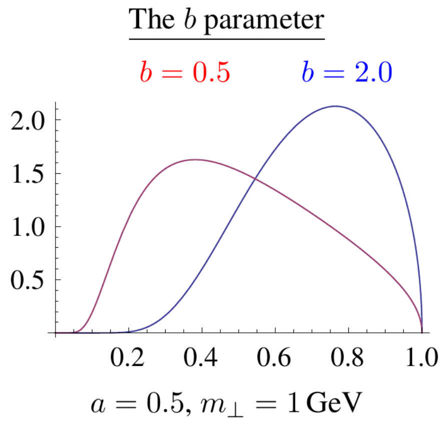
p_z

The momentum fraction z of each fragmenting hadron is sampled according to the

Lund fragmentation function

$$f(z) \propto \frac{(1-z)^a}{z} \exp\left(\frac{-bm_\perp^2}{z}\right), \qquad z = \frac{p_z + E_h}{2E}$$



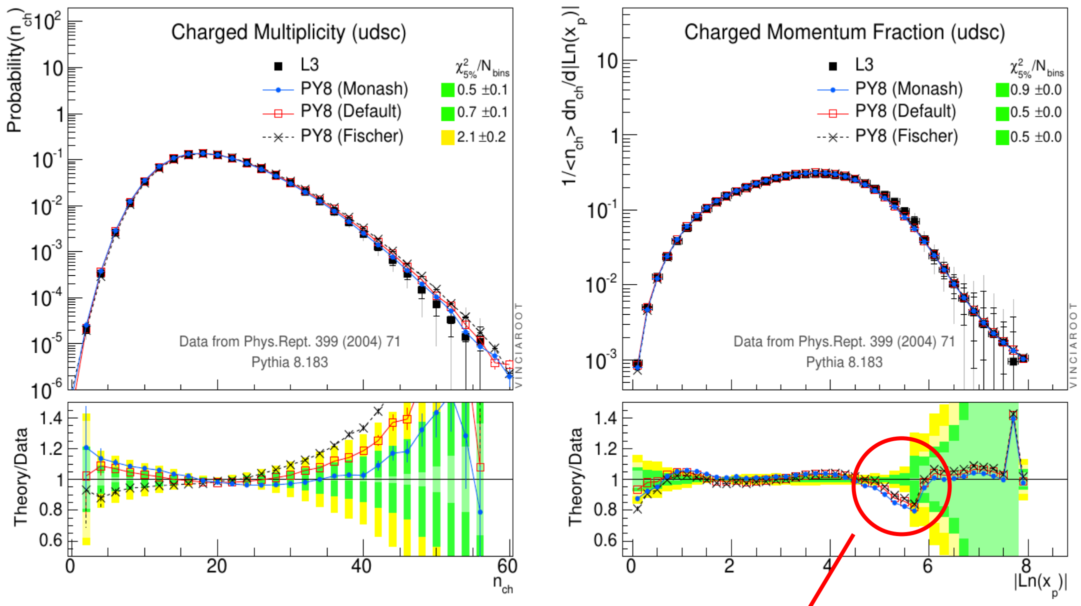


Tuning

Model parameters must be tuned on experimental observables.

Hadronization parameters a,b, and σ_{pT} are tuned using event shape variables + multiplicity and momentum fraction distributions from LEP (L3).

- **Unavoidable discrepancies with data due largely to highly correlated parameter space**



Monash Tune (a,b,σ_{perp}) 1404.5630

Cannot be improved by retuning

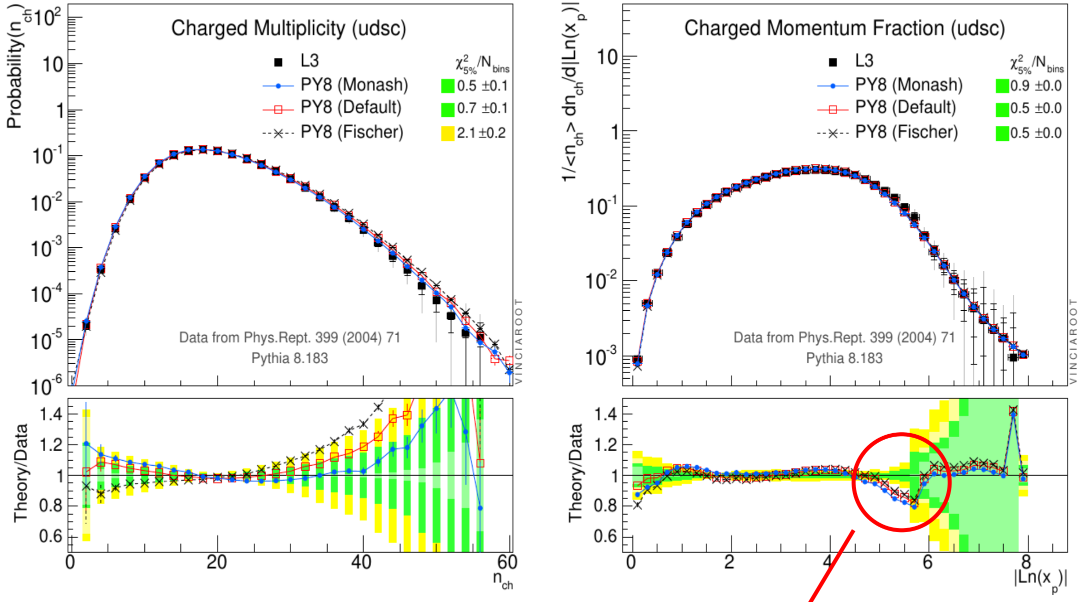
Tuning

Promotes two lines of invesitgation:

1. A search for more hadronization sensitive observables that will allow for independent parameter tuning

2. An implementation of a new model of hadronization the incorporates additional degrees of freedom such that the model can accommodate all data.

- **Unavoidable discrepancies with data due largely to highly correlated parameter space**



Monash Tune (a, b, σ_{perp}) 1404.5630

Cannot be improved by retuning

How to improve agreement with data: two approaches

- Improve model
  - MPIs, rope hadronization, transverse mass supression, flavor asymmetries, hadronic rescattering, multiscale models (string → hydrodynamical), flavor selector, etc.
  - Utilize techniques from gauge-gravity duality

  Hard to come up with mathematically precise model without established calculational techniques

How to improve agreement with data: two approaches

- Improve model
  - MPIs, rope hadronization, transverse mass supression, flavor asymmetries, hadronic rescattering, multiscale models (string → hydrodynamical), flavor selector, etc.
  - Utilize techniques from gauge-gravity duality

  Hard to come up with mathematically precise model without established calculational techniques

- Tend towards model independendence
  - Sample directly from global distributions

  Non-universal and extremely difficult to convert into representative particle flow data

How to improve agreement with data: two* approaches

- Improve model
  - MPIs, rope hadronization, transverse mass supression, flavor asymmetries, hadronic rescattering, multiscale models (string → hydrodynamical), flavor selector, etc.
  - Utilize techniques from gauge-gravity duality

  Hard to come up with mathematically precise model without established calculational techniques

- Tend towards model independendence
  - Sample directly from global distributions

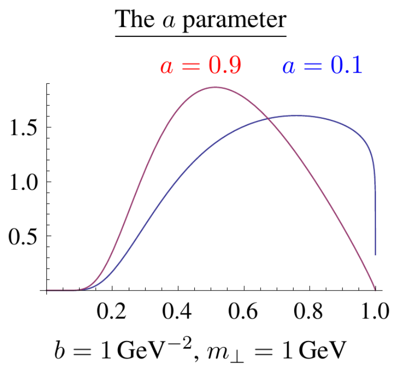
  Non-universal and extremely difficult to convert into representative particle flow data

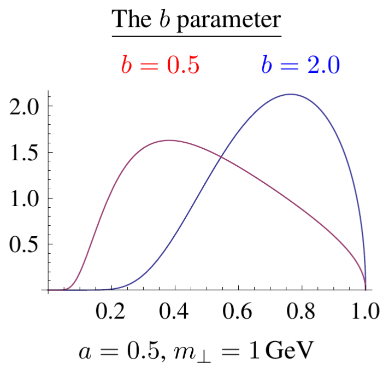
*** or some combination of both**

Hybrid data-driven approach

The phenomenological models of hadronization already give an acceptable description of a large amount of data.

Hybrid approach: Keep the underlying paradigm i.e. strings but modify the microscopic kinematics to accommodate the discrepant global experimental observables.



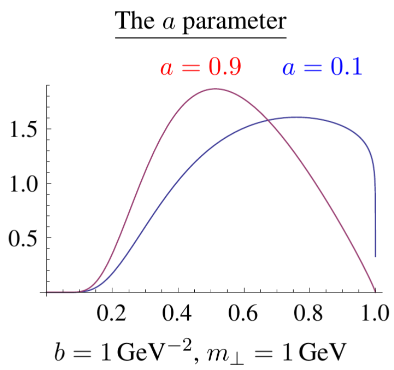


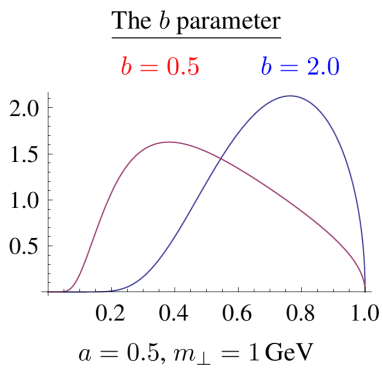
+ ε

Hybrid data-driven approach

The phenomenological models of hadronization already give an acceptable description of a large amount of data.

Hybrid approach: Keep the underlying paradigm i.e. strings but modify the microscopic kinematics to accommodate the discrepant global experimental observables.





+ ε

Machine learning (ML) offers a nice framework to tackle this problem

Steps to data-driven model

A) Implement a ML based model that can fully reproduce Pythia output

Steps to data-driven model

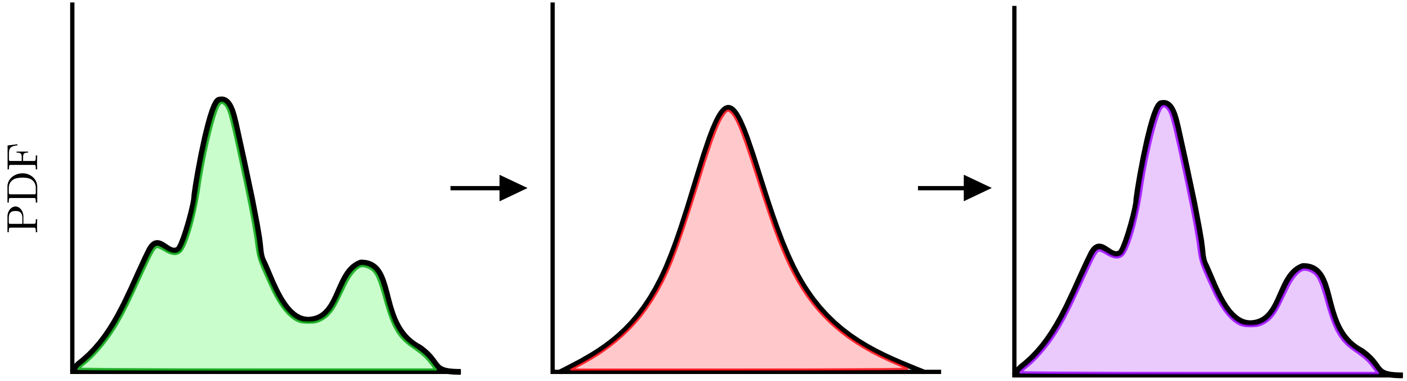
A) Implement a ML based model that can fully reproduce Pythia output

B) Modify the microscopic dynamics (kinematics) of the model to accomodate any discrepancies with macroscopic observable distributions

A) A ML model of hadronization

To make any headway we need a tool which will allow us to efficiently sample probability distributions whose analytic form is unknown.

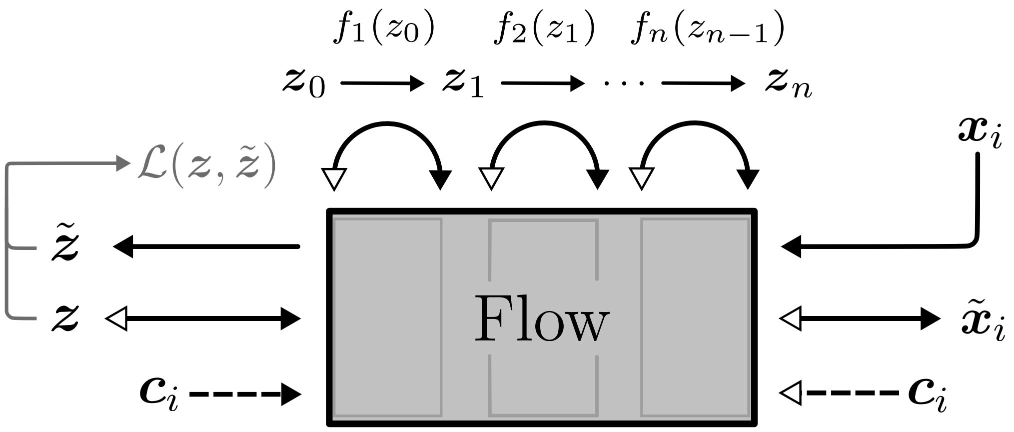
Generative machine learning algorithms are the perfect tool



Invertible nerual networks (INN)

a.k.a normalizing flow

Learn a composition of n independent bijective transformations that relate a probability distribution $p_Z(z)$ on latent space Z to the target distribution $p_X(x)$ on target space X.



The probability distribution for the random variable x = f(z) is given by

$$p_X^f(\boldsymbol{x}) = p_Z(\boldsymbol{z})|\det J_f(\boldsymbol{z})|^{-1}$$

$$J_f = \partial f / \partial x$$

For n iterative transformations:

$$p_X^F(\boldsymbol{x}) = p_Z(\boldsymbol{z}_0) \prod_{i=1}^{n} |\det J_{f_i}(\boldsymbol{z}_{i-1})|^{-1}$$

Invertible nerual networks (INN)

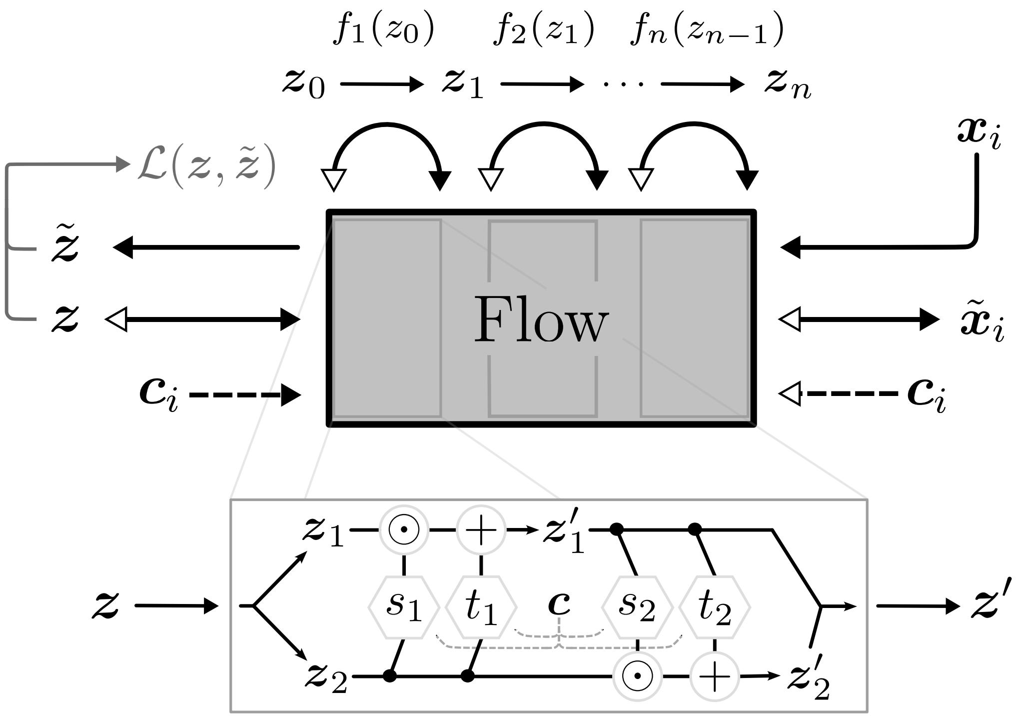
a.k.a normalizing flow

Invertible Real NVP transformations:

$$z'_1 = z_1 \odot \exp\left(s_1\left(z_2\right)\right) + t_1(z_2),$$
$$z'_2 = z_2 \odot \exp\left(s_2\left(z'_1\right)\right) + t_2(z'_1),$$

Scale transform

Translation transform



Invertible nerual networks (INN)

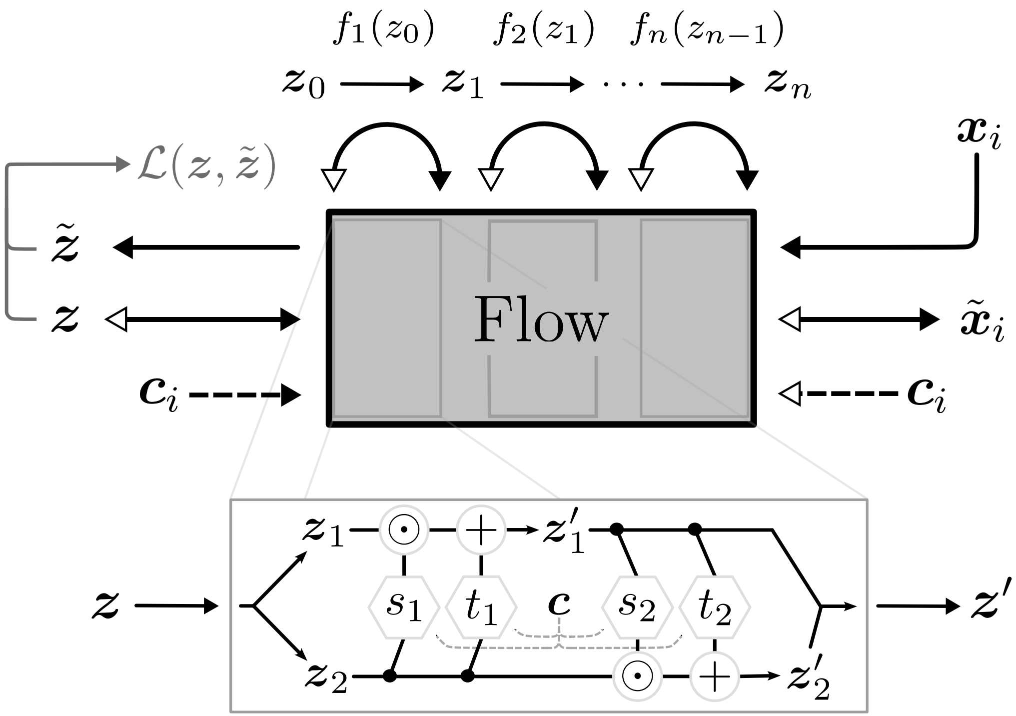
a.k.a normalizing flow

Invertible Real NVP transformations:

$$z_1' = z_1 \odot \exp\big(s_1(z_2)\big) + t_1(z_2),$$
$$z_2' = z_2 \odot \exp\big(s_2(z_1')\big) + t_2(z_1'),$$

Inverse:

$$z_2 = \big(z_2' - t(z_1')\big) \odot \exp\big(-s(z_1')\big),$$
$$z_1 = \big(z_1' - t(z_2)\big) \odot \exp\big(-s(z_2)\big),$$



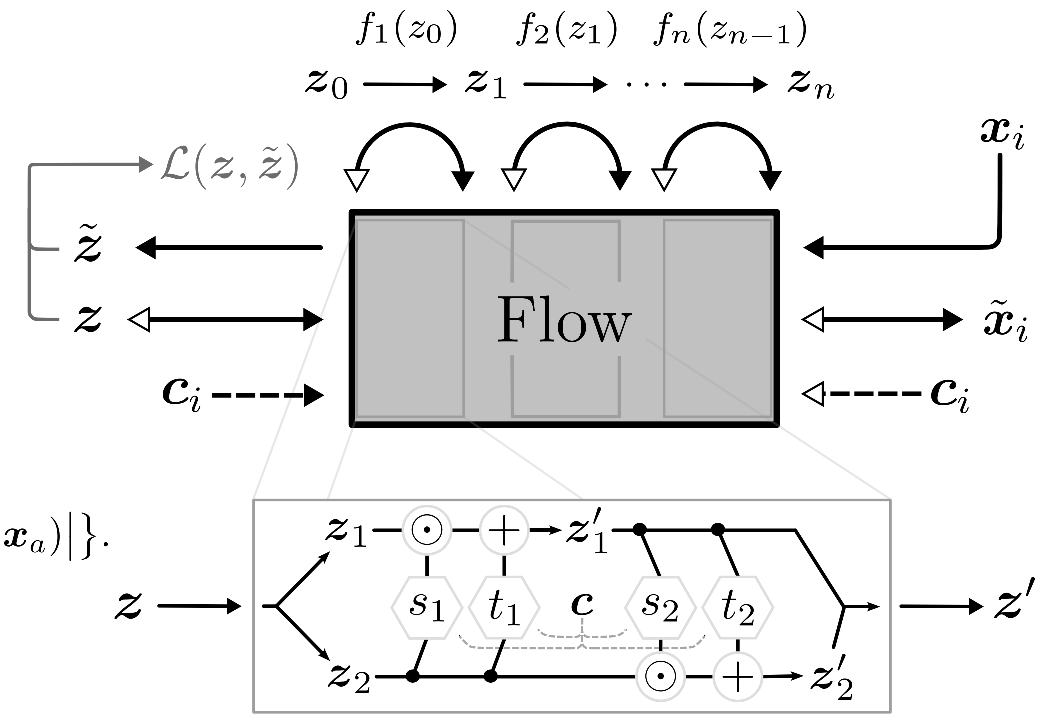
Invertible nerual networks (INN)

a.k.a normalizing flow

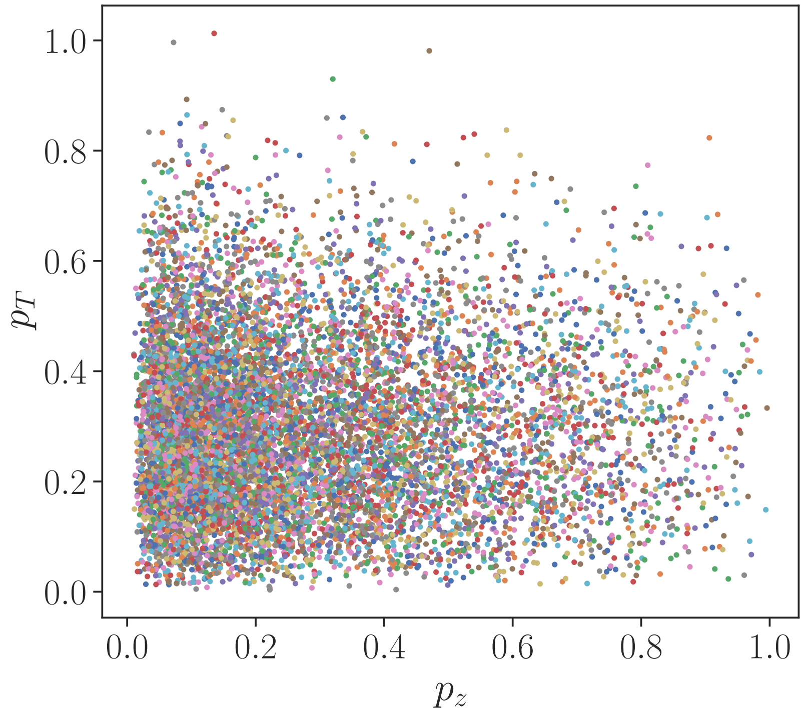
$$p_X^F(\boldsymbol{x}) = p_Z(\boldsymbol{z}_0)\prod_{i=1}^{n}\left|\det J_{f_i}(\boldsymbol{z}_{i-1})\right|^{-1}$$

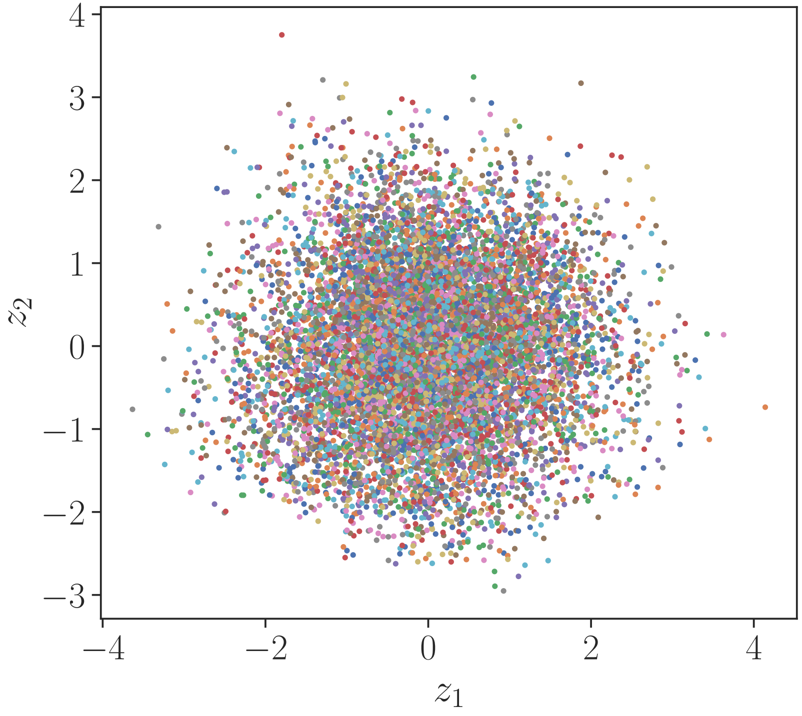
Train with the negative log likelihood:

$$\mathcal{L}_{\rm NF} = -\sum_{a=1}^{N}\log p_X^F(\boldsymbol{x}_a;\boldsymbol{\theta},\boldsymbol{c}_a)$$
$$= \sum_{a=1}^{N}\left\{-\log p_Z\left(F^{-1}(\boldsymbol{x}_a;\boldsymbol{\theta},\boldsymbol{c}_a)\right) + \log\left|\det J_{F^{-1}}(\boldsymbol{x}_a)\right|\right\}.$$

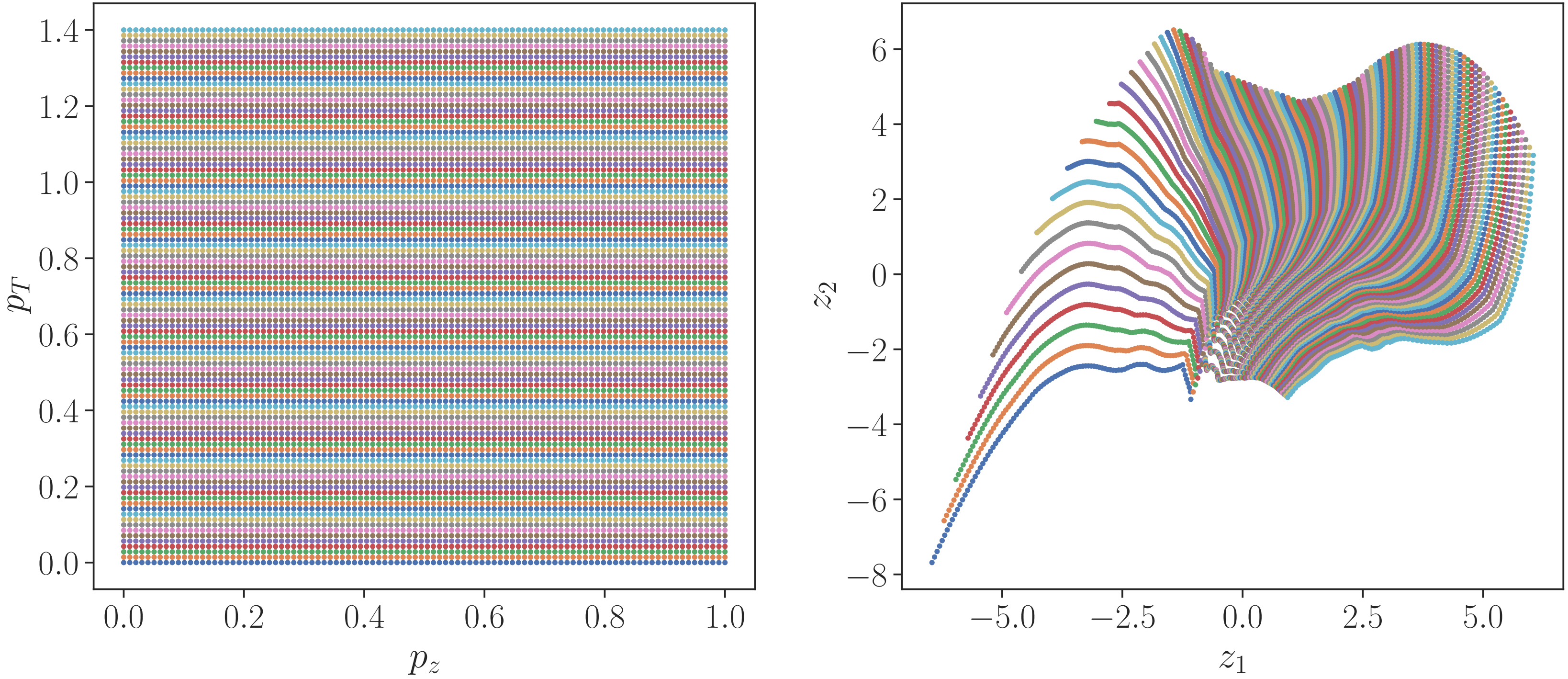


INN learned mapping



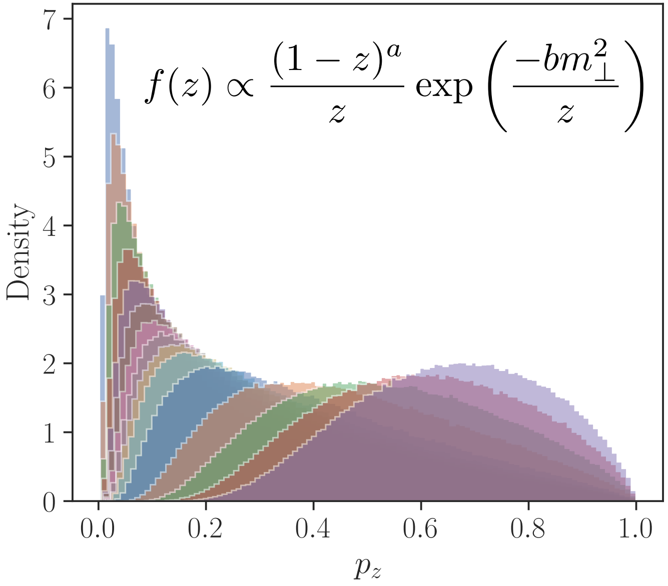


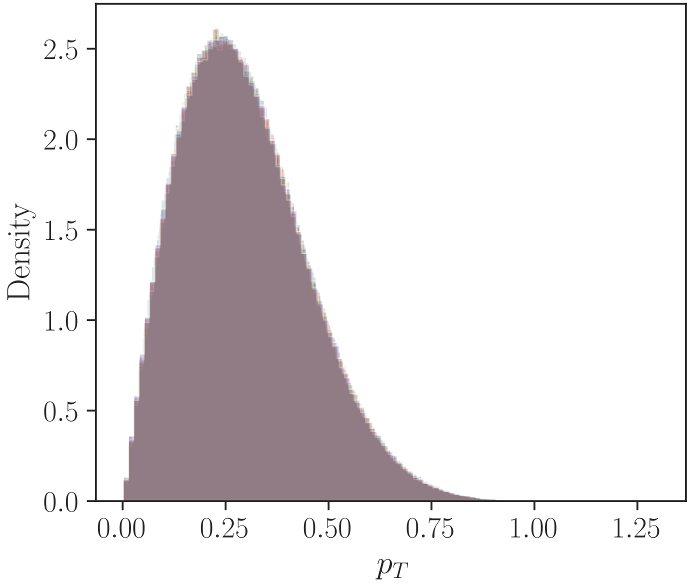
INN learned mapping



Training data

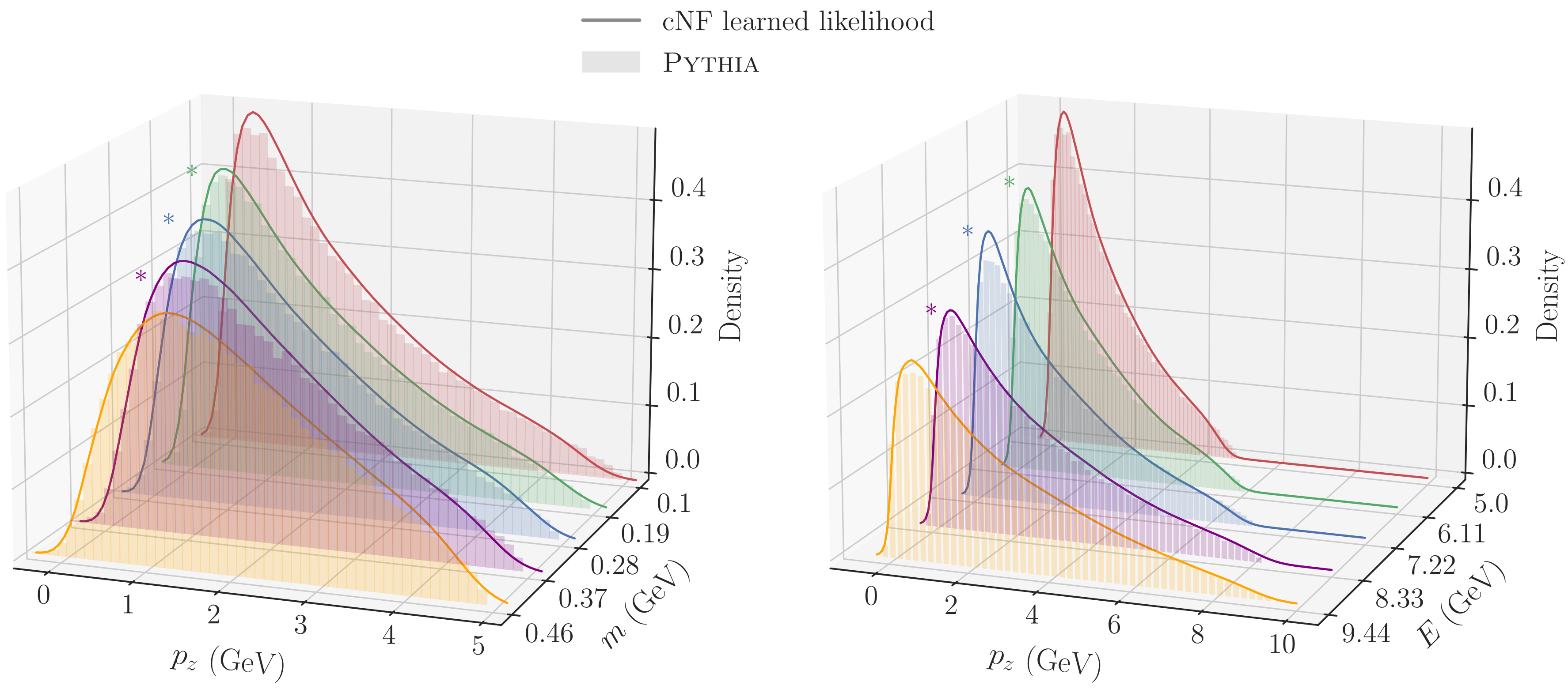
The implementation of full hadronization event using single emission kinematics requires an independent p_T sampling followed by a p_T-dependent sampling of p_z sampling (due to the dependence of f(z) of transverse mass).



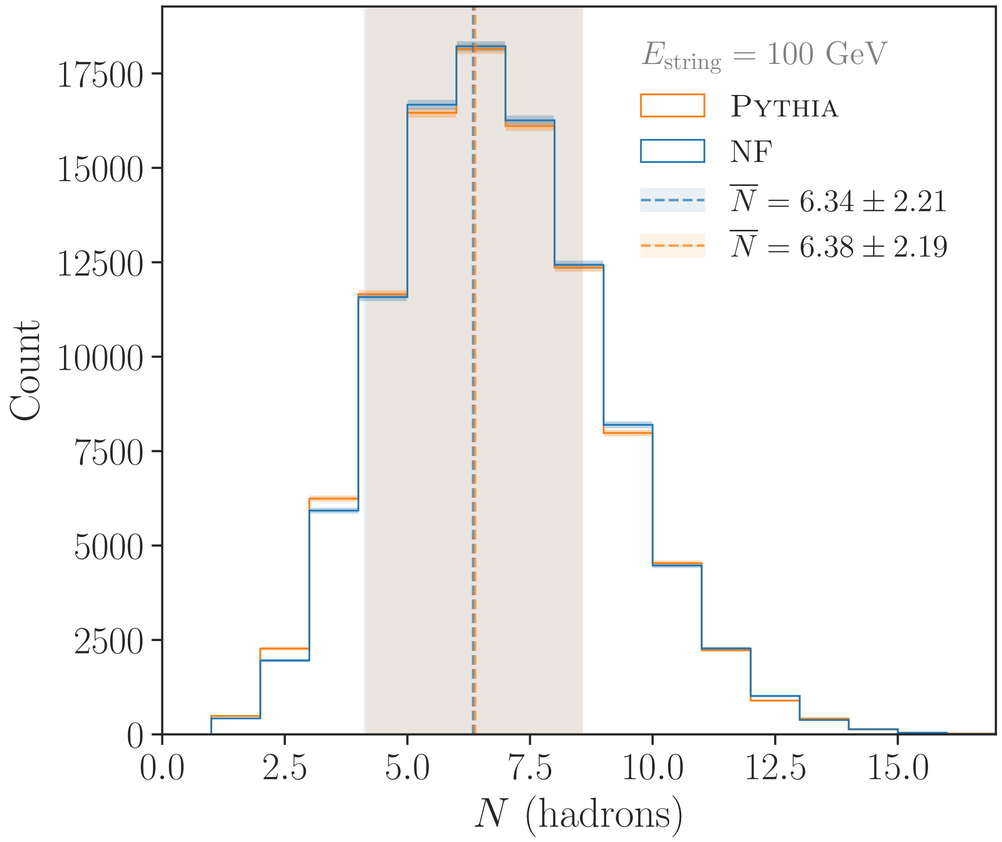


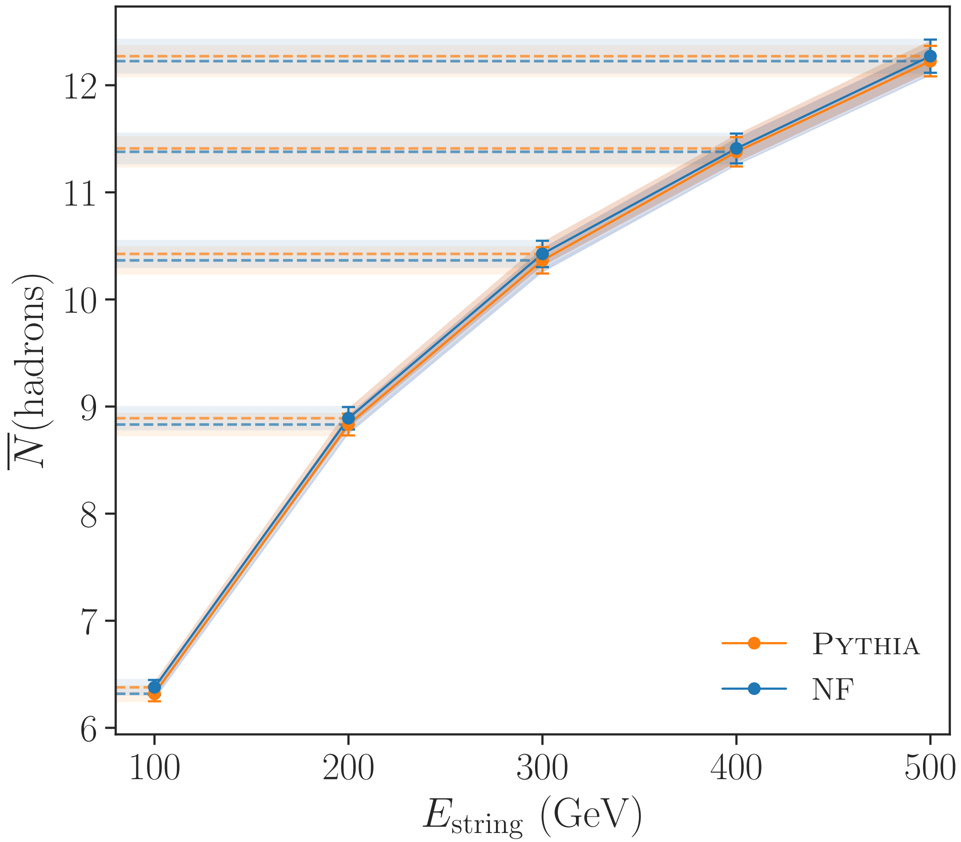
Generate Pythia $q\bar{q} \to h$ events, first with no p_T kicks at different values of the hadron mass, record p_z. Generate events again, with kicks turned on, record p_T.

Validation: single emission kinematics



Validation: global observable (hadron multiplicity N)





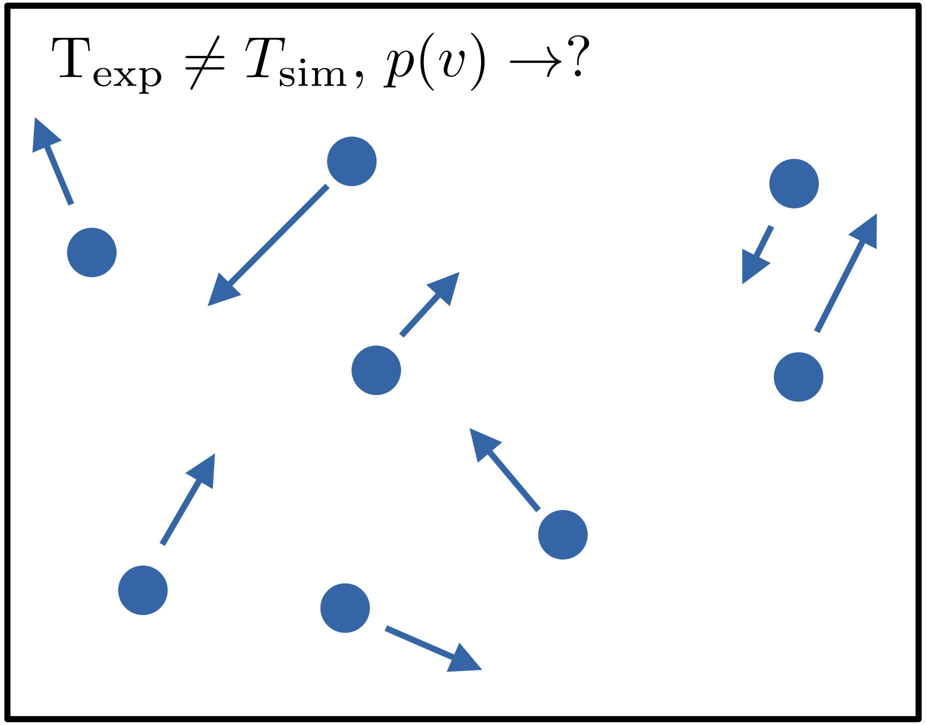
B) Training microscopic dynamics from macroscopic observables

Task: Given a discrepant macroscopic observable between simulation and experiment, find a way to modify the microscopic dynamics of the simulation to accommodate the observable.

B) Training microscopic dynamics from macroscopic observables

Task: Given a discrepant macroscopic observable between simulation and experiment, find a way to modify the microscopic dynamics of the simulation to accommodate the observable.

Akin to...



Microscopic Alterations Generated from IR Collections (MAGIC)

Two step training paradigm:

1) Train a base model

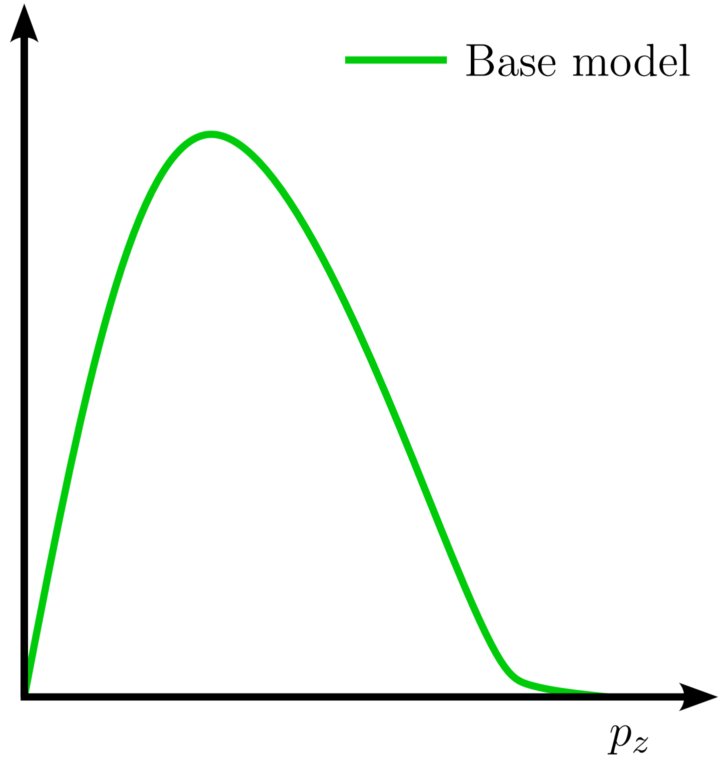
    a) Start from where we left off in step A

Microscopic Alterations Generated from IR Collections (MAGIC)

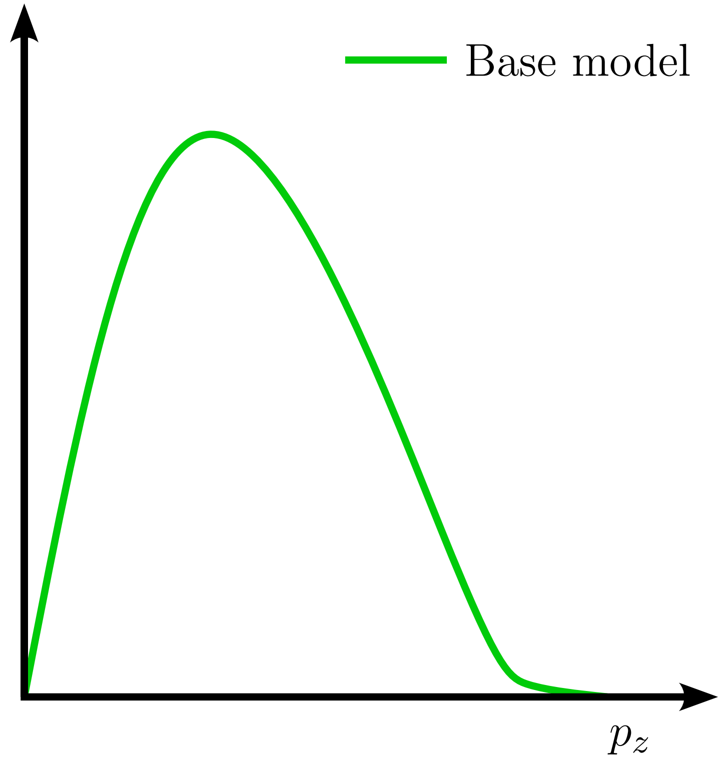
Two step training paradigm:

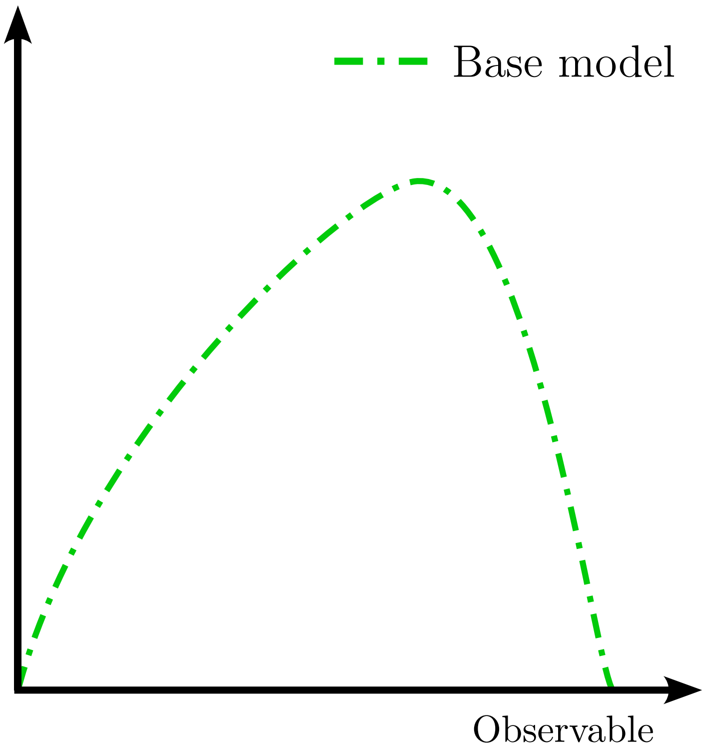
1) Train a base model
   a) Start from where we left off in step A
2) Fine tune
   a) Perturb the p_z and p_T single emission distributions, generate and compare global observable distributions, iterate

MAGIC: fine tuning

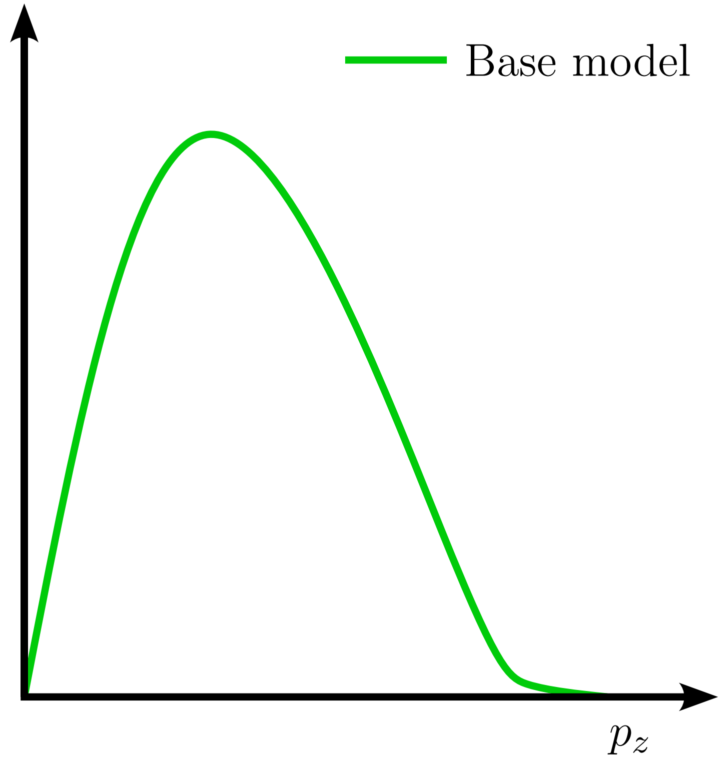


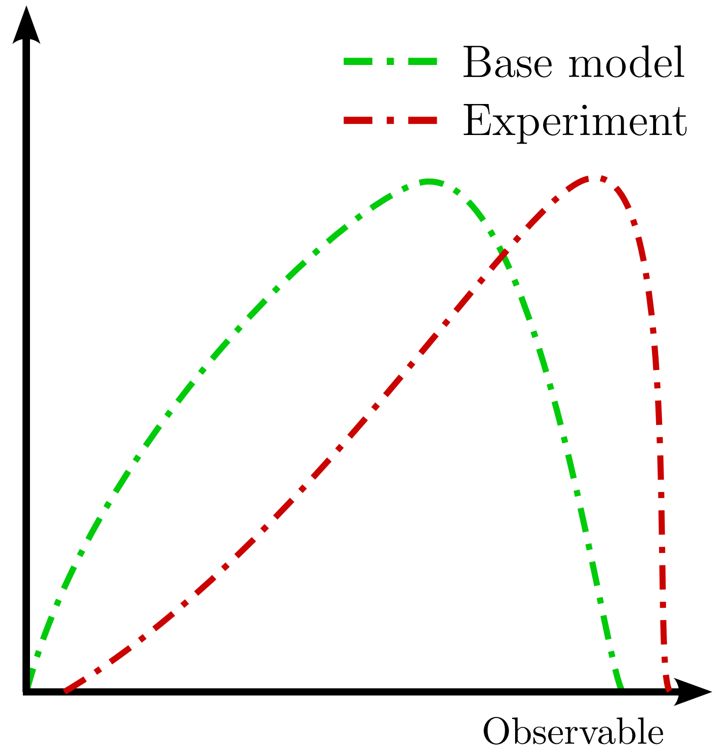
MAGIC: fine tuning



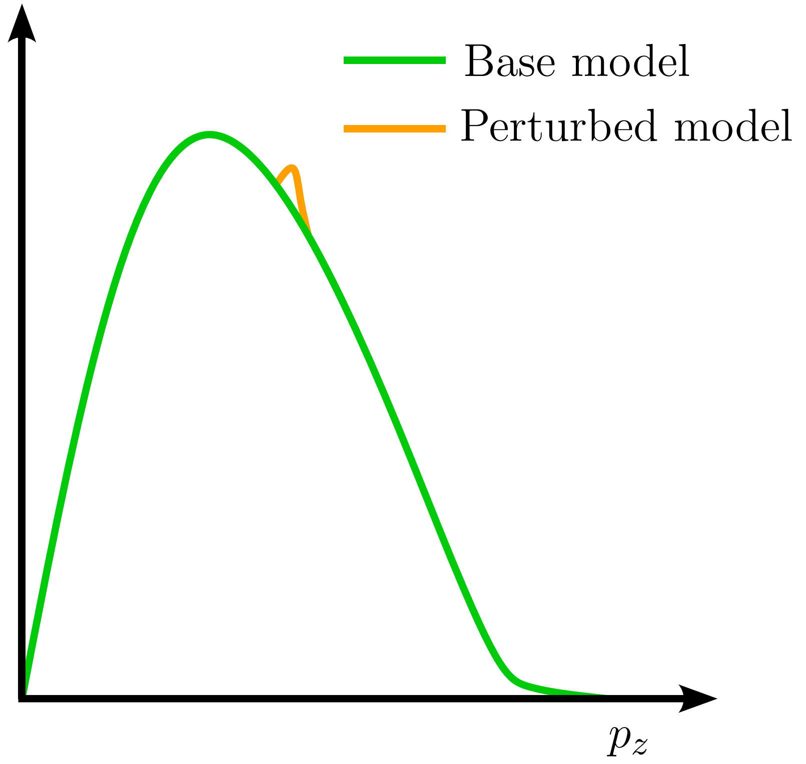


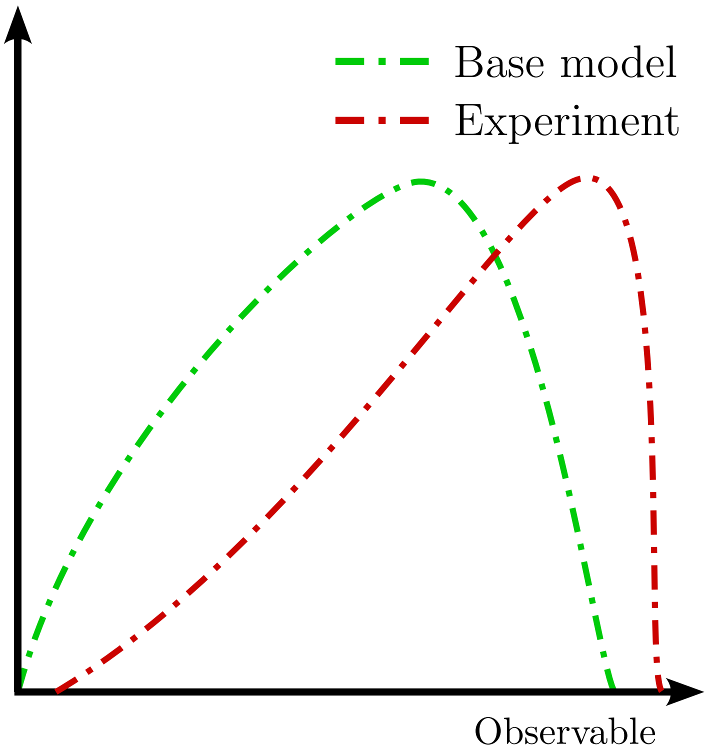
MAGIC: fine tuning



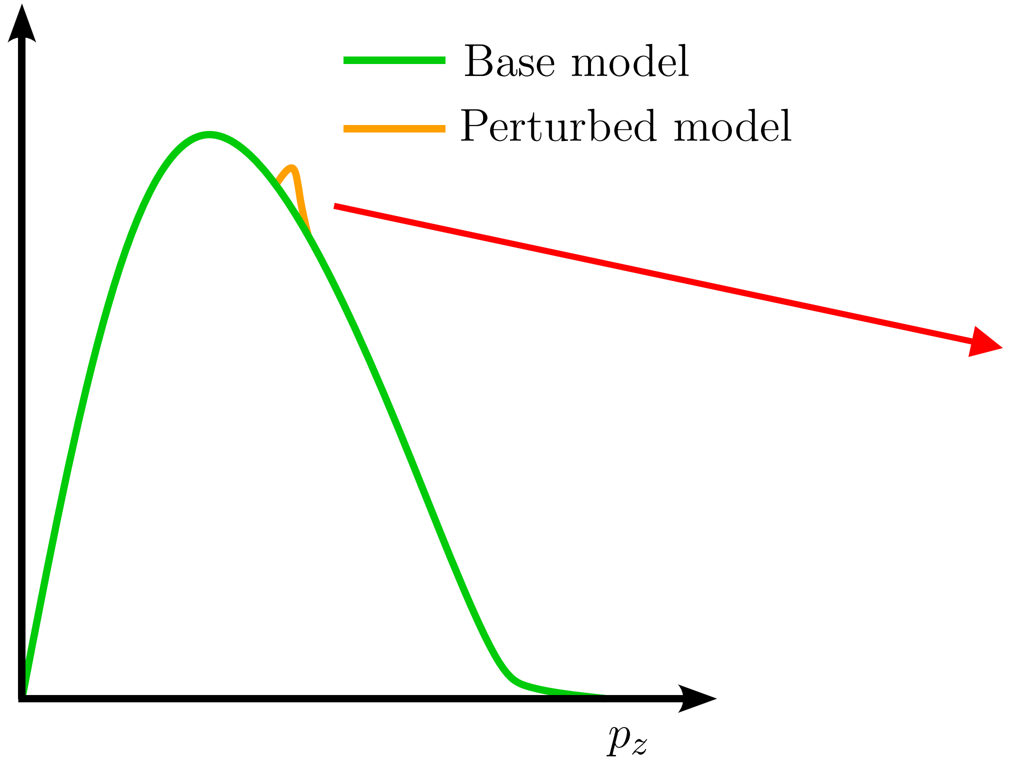


MAGIC: fine tuning



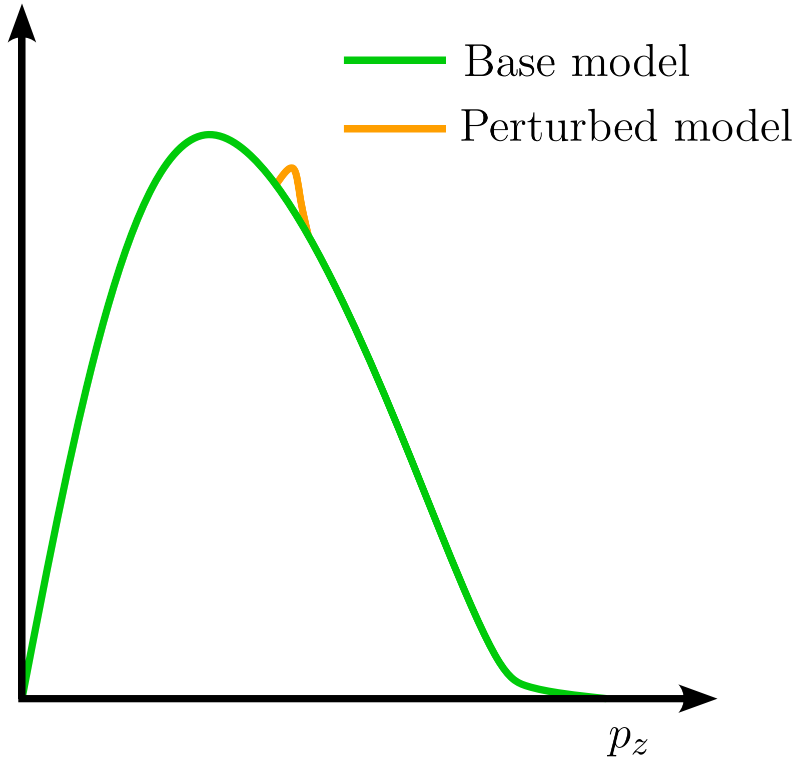


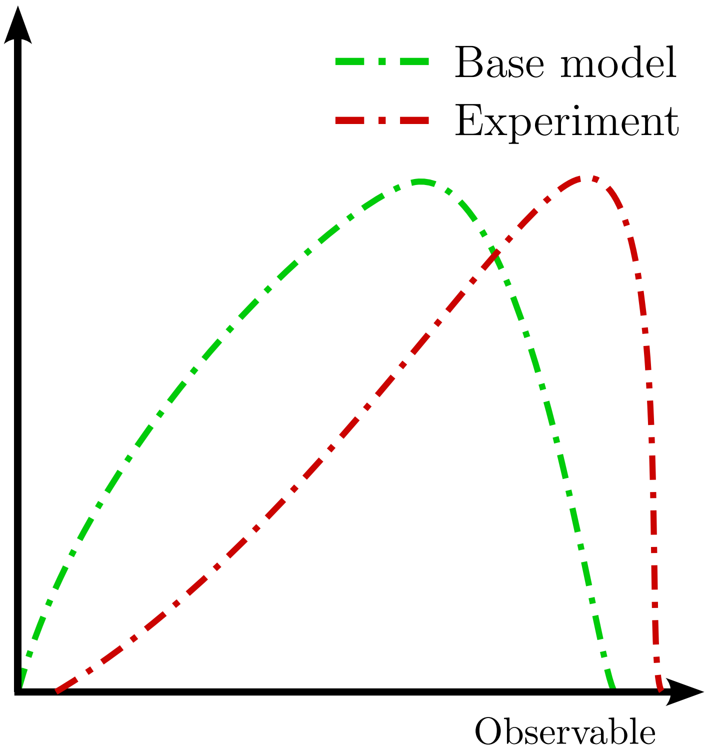
MAGIC: fine tuning



This perturbation can, in principle, be linked to a change in the parameters of the INN affine coupling blocks i.e. the weights in one of the scale or translation MLPs

MAGIC: fine tuning





MAGIC: fine tuning

A road block:

Generally speaking at this point we would need to re-simulate events in order to understand how the perturbation effects the macroscopic observable distribution.

Re-simulating at each model update is computationally infeasible.

MAGIC: fine tuning

A road block:

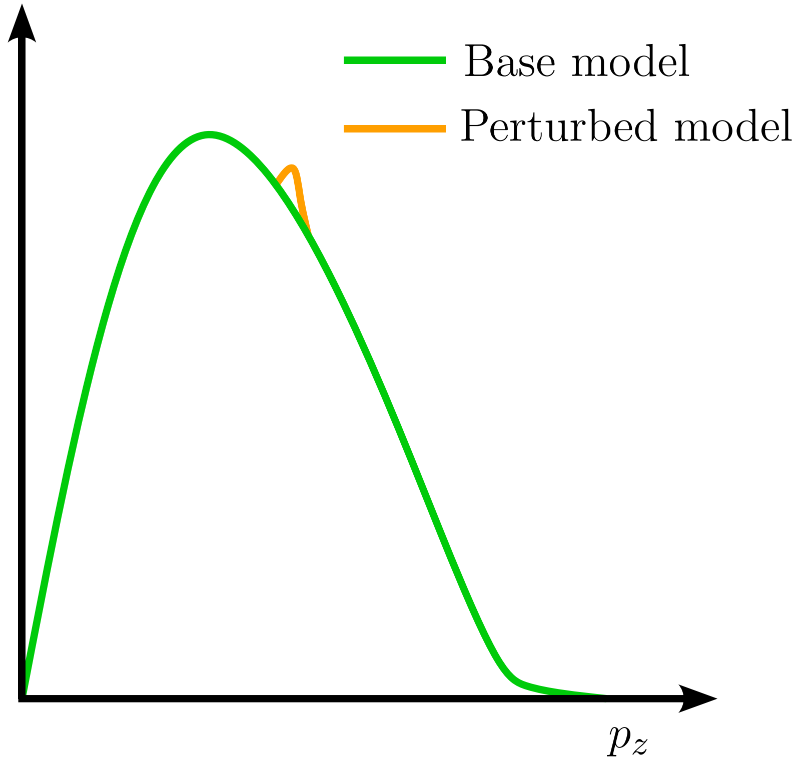
Generally speaking at this point we would need to re-simulate events in order to understand how the perturbation effects the macroscopic observable distribution.

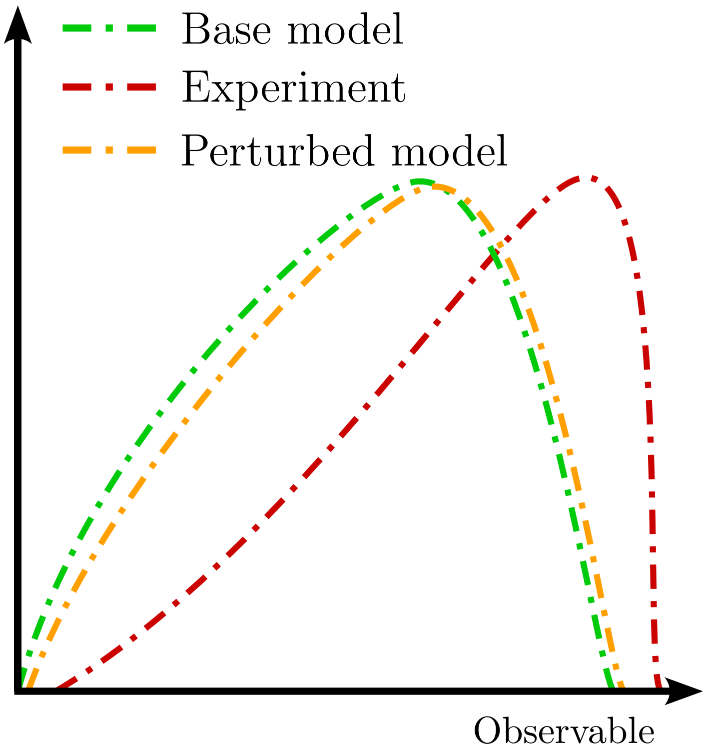
Re-simulating at each model update is computationally infeasible.

The solution:

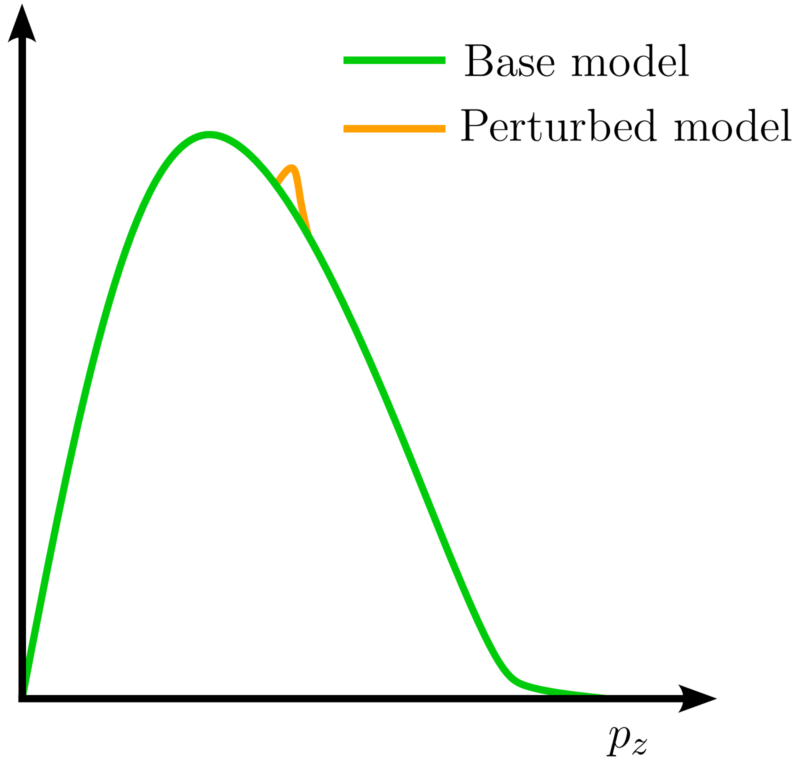
Use the NF likelihood to reweigh events!

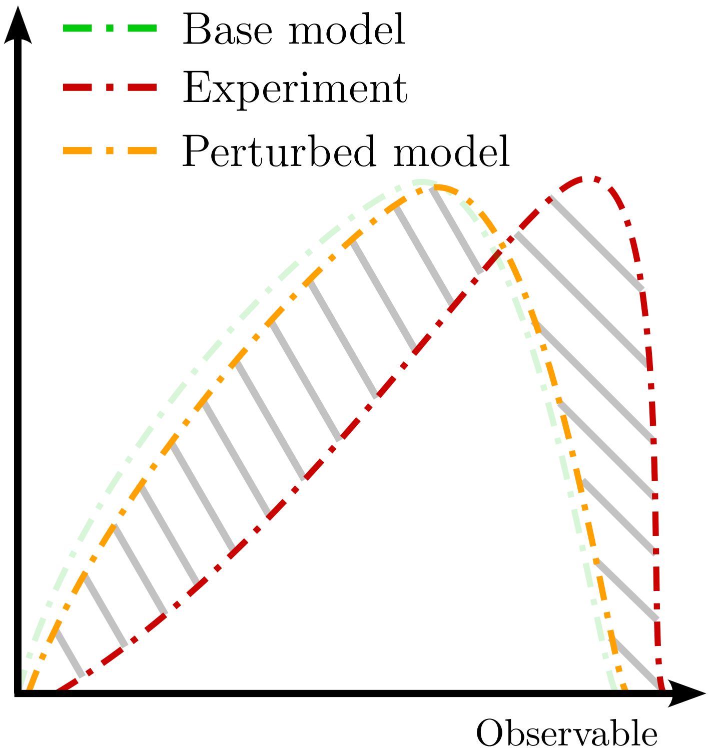
MAGIC: fine tuning



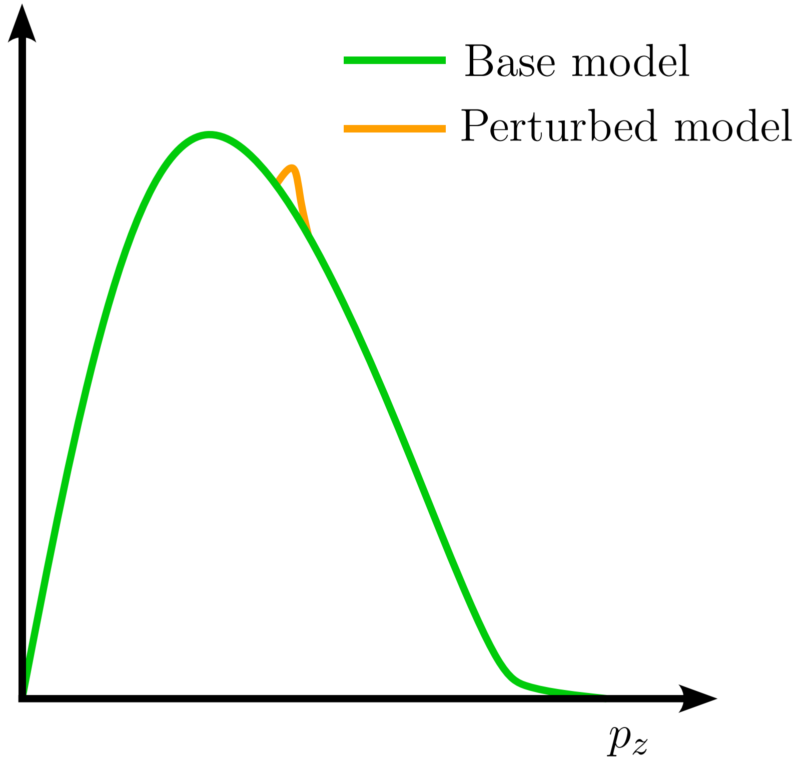


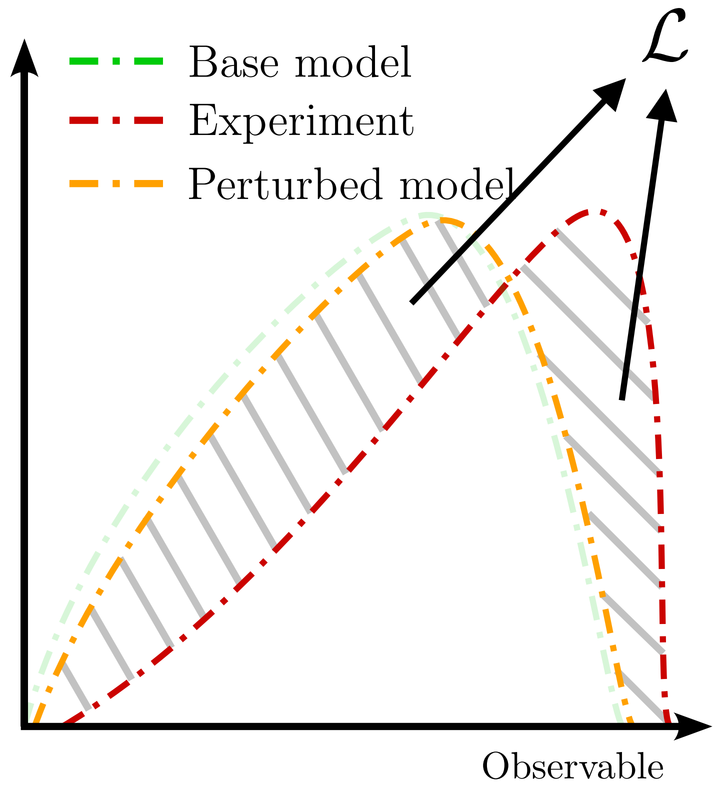
MAGIC: fine tuning



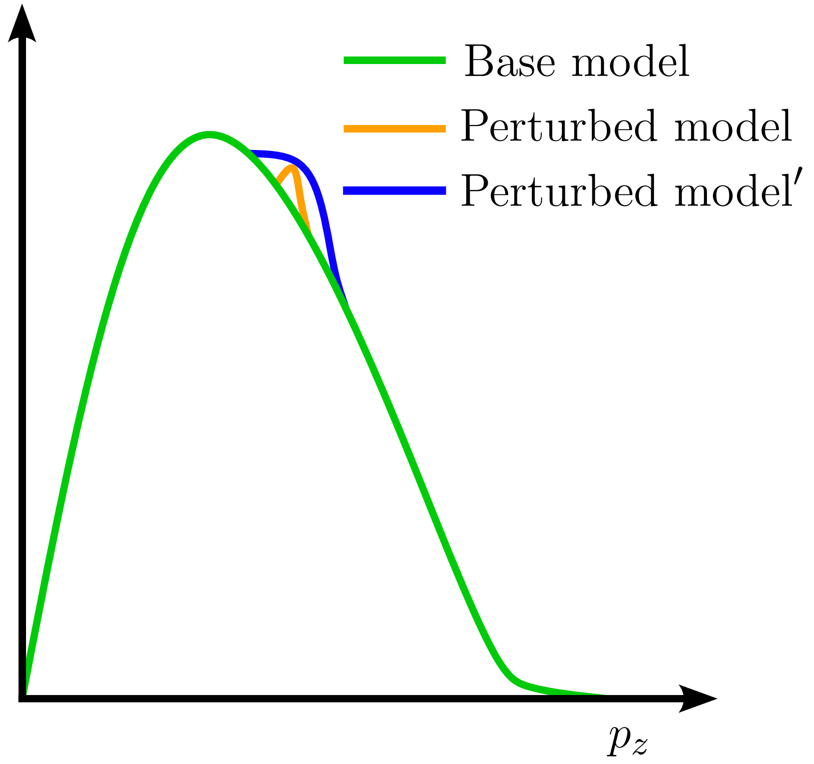


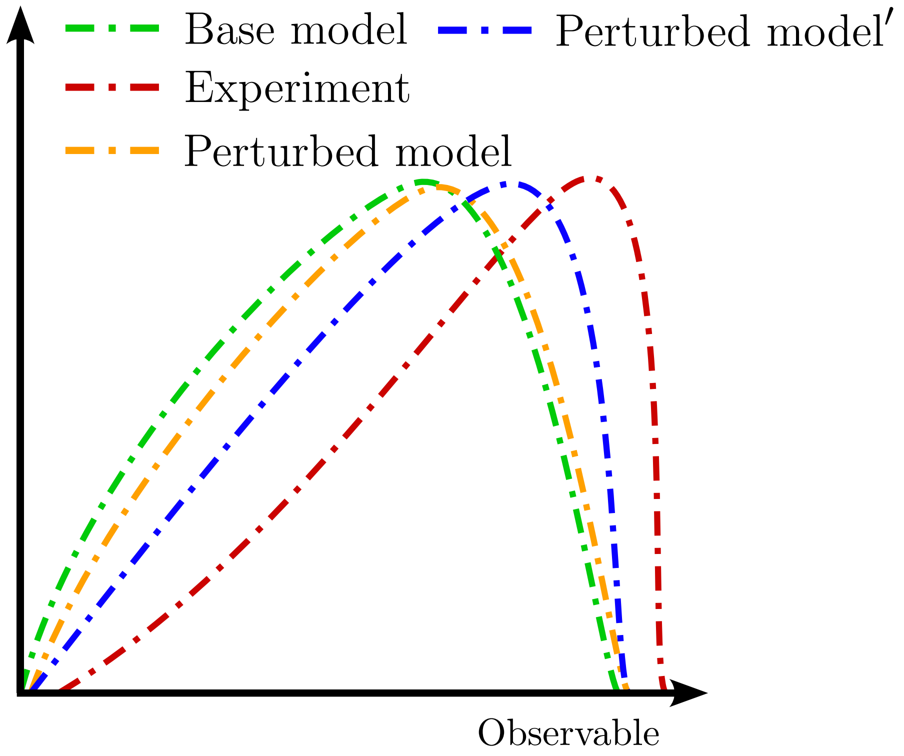
MAGIC: fine tuning



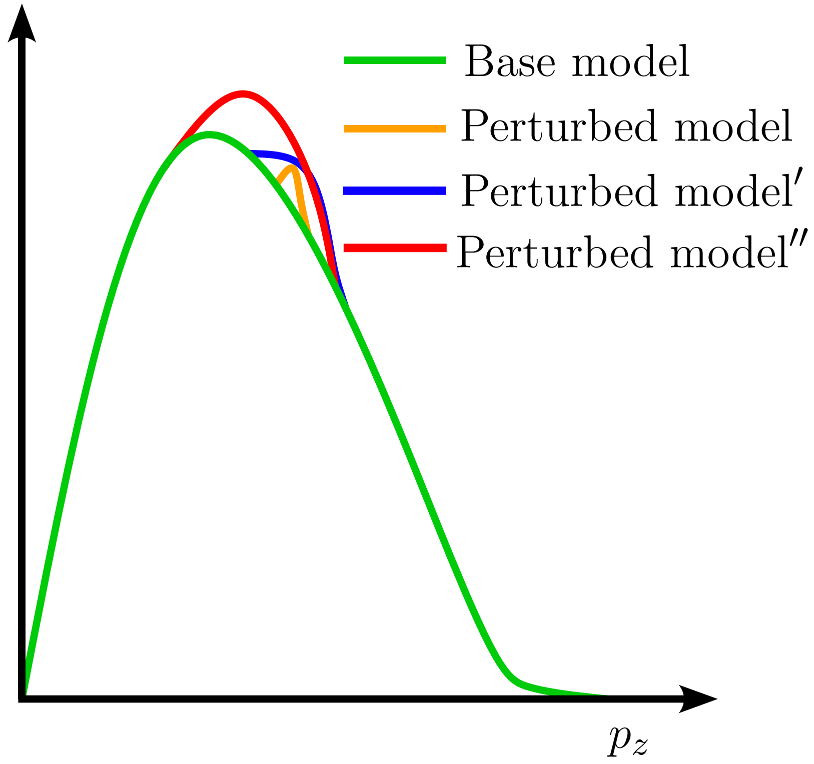


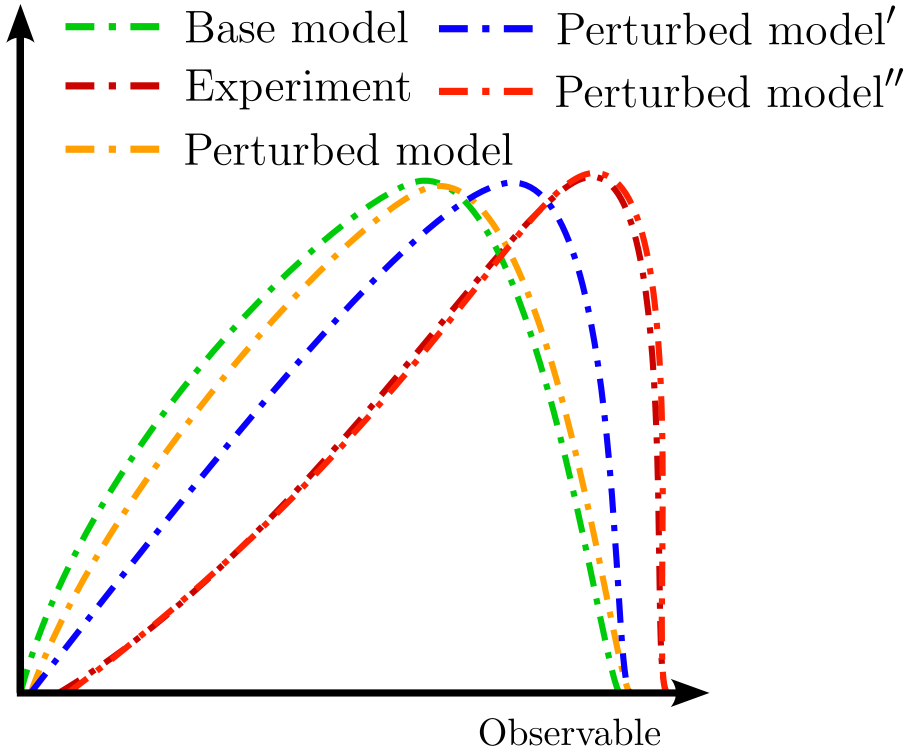
MAGIC: fine tuning





MAGIC: fine tuning





Toy example

Three ingredients:

1. Event level observables Y^{sim} from base model simulated hadronization events
2. (p_z, p_T) values for all fragmentations in each simulated hadronization event
3. Event level observables Y^{exp} from experimental distributions

Toy example

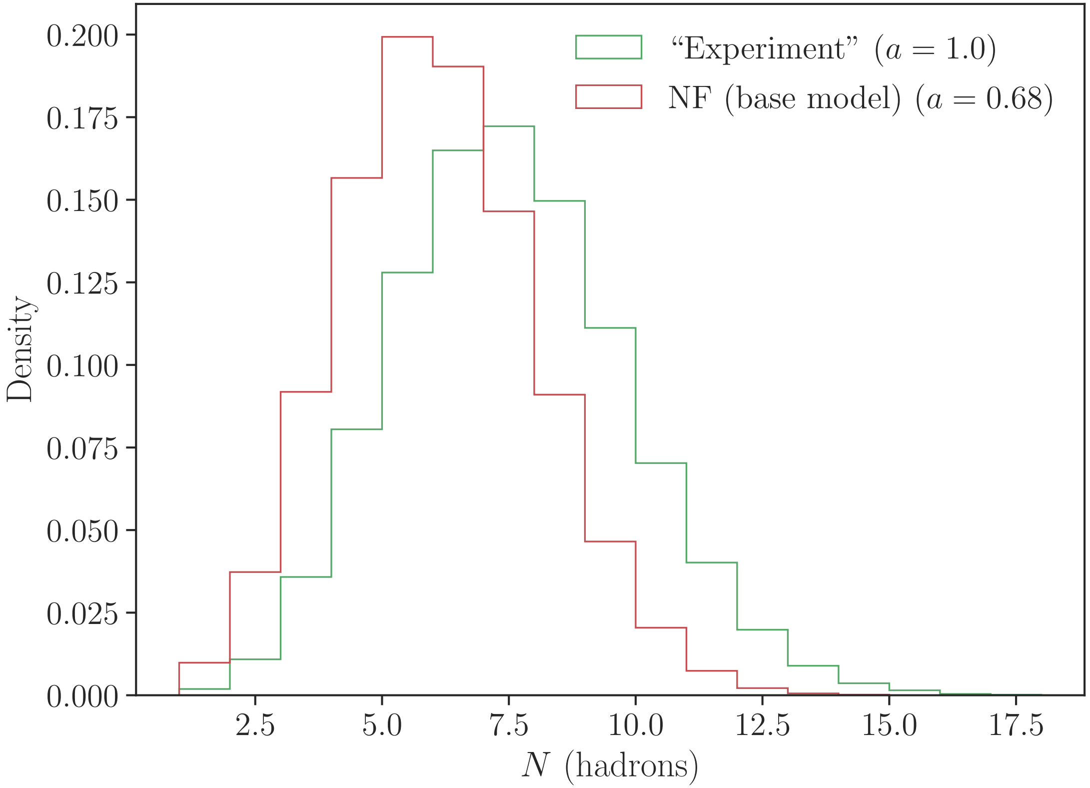
Use hadron multiplicity N as global event-level observable

$$\boldsymbol{X} = \begin{pmatrix} \left[\{p_z^{h_1}, p_T^{h_1}\}, \{p_z^{h_2}, p_T^{h_2}\}, \{p_z^{h_3}, p_T^{h_3}\}\right]_1 \\ \left[\{p_z^{h_1}, p_T^{h_1}\}, \{p_z^{h_2}, p_T^{h_2}\}, \{p_z^{h_3}, p_T^{h_3}\}, \{p_z^{h_4}, p_T^{h_4}\}, \{p_z^{h_5}, p_T^{h_5}\}\right]_2 \\ \vdots \\ \left[\{p_z^{h_1}, p_T^{h_1}\}, \{p_z^{h_2}, p_T^{h_2}\}\right]_n \end{pmatrix}, \quad \boldsymbol{Y}^{\text{sim}} = \begin{pmatrix} N_1 & = 3 \\ N_2 & = 5 \\ & \vdots \\ N_n & = 2 \end{pmatrix}, \quad \boldsymbol{Y}^{\text{exp}} = \begin{pmatrix} N_1 & = 7 \\ N_2 & = 3 \\ & \vdots \\ N_n & = 2 \end{pmatrix}$$

Event weights: $$\boldsymbol{w} = \begin{pmatrix} \prod_{i=1}^{N_1} w_i \\ \prod_{j=1}^{N_2} w_j \\ \vdots \\ \prod_{k=1}^{N_n} w_k \end{pmatrix} \quad \text{where} \quad w_i = \frac{p_X^{F'}(p_z^{h_i}, p_T^{h_i})}{p_X^{F}(p_z^{h_i}, p_T^{h_i})},$$

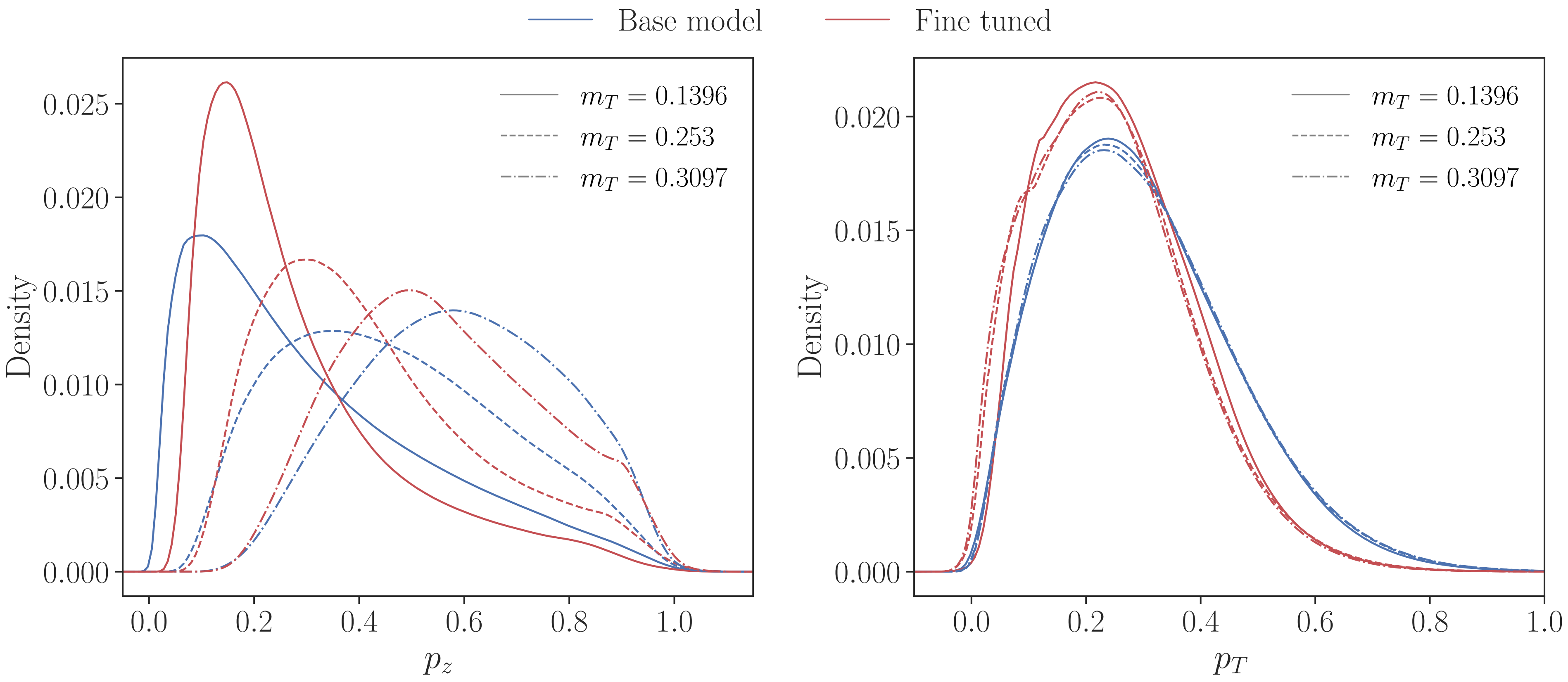
Loss (Earth mover's distance): $\mathcal{L}_{\text{EMD}}(\boldsymbol{Y}^{\text{sim}}, \boldsymbol{w}, \boldsymbol{Y}^{\text{exp}})$

Toy example



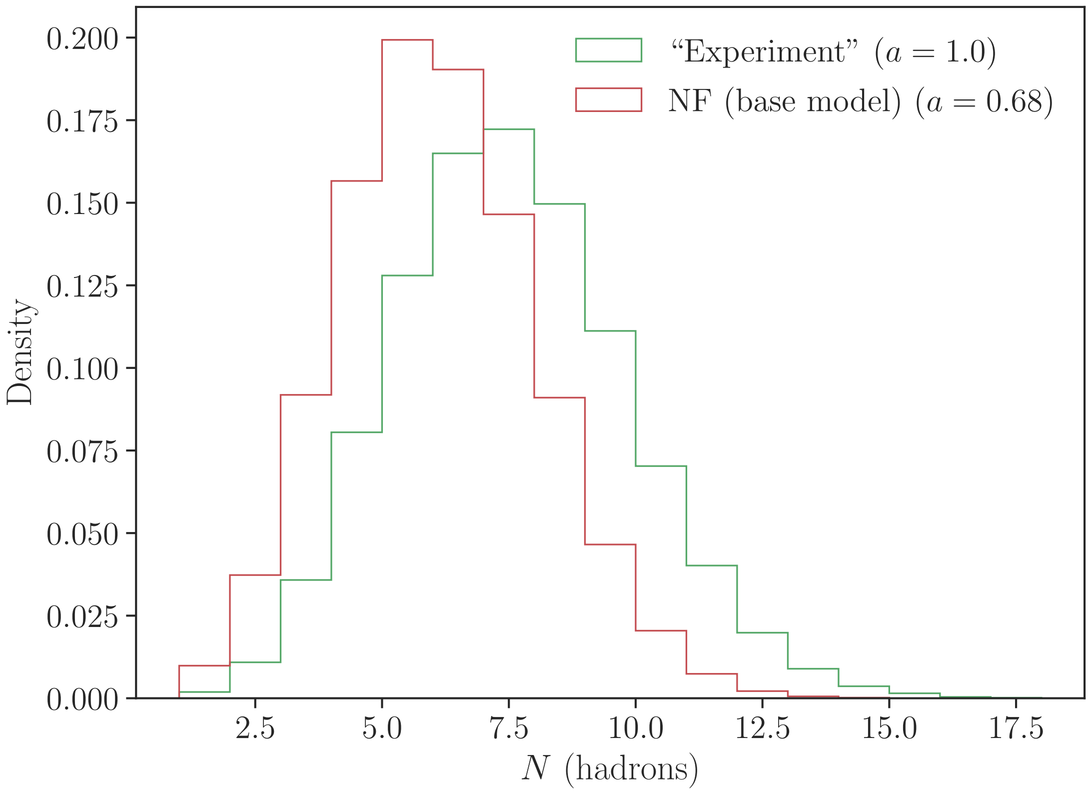
“Experiment” ($a = 1.0$)
NF (base model) ($a = 0.68$)
Density
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
0.200
2.5
5.0
7.5
10.0
12.5
15.0
17.5
N (hadrons)

Toy example



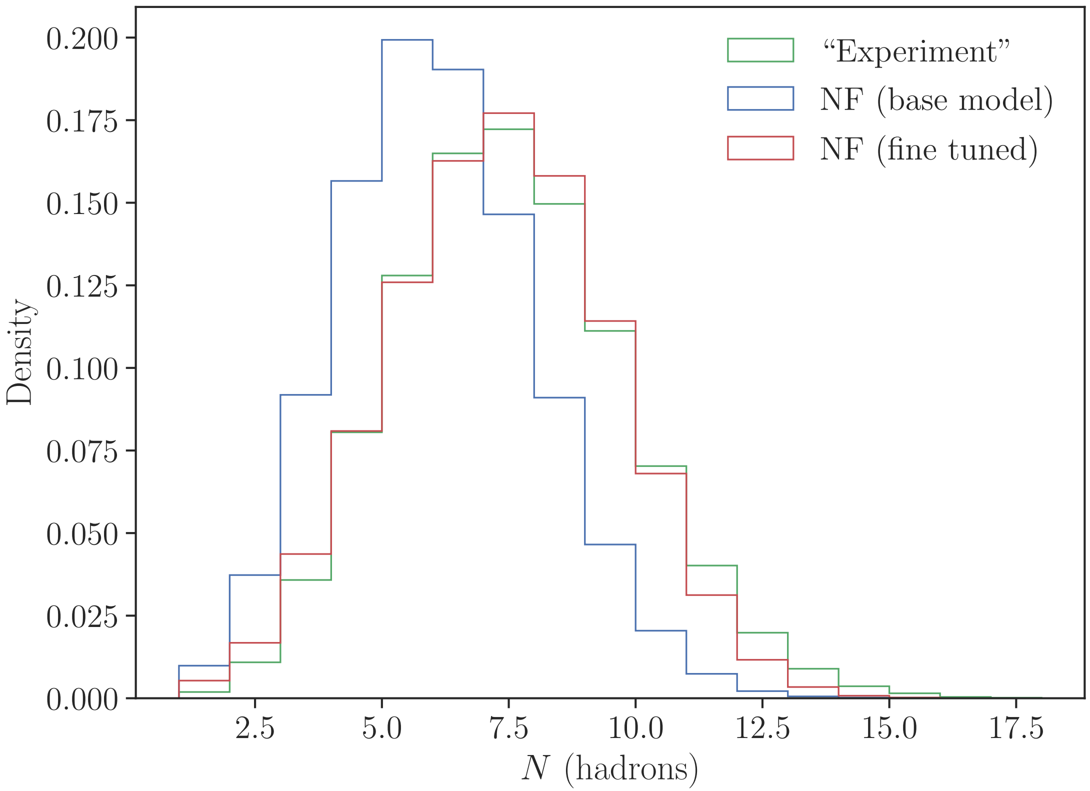
Base model
Fine tuned
$m_T = 0.1396$
$m_T = 0.253$
$m_T = 0.3097$
Density
p_z
$m_T = 0.1396$
$m_T = 0.253$
$m_T = 0.3097$
Density
p_T

Toy example



"Experiment" ($a = 1.0$)
NF (base model) ($a = 0.68$)
Density
0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
2.5
5.0
7.5
10.0
12.5
15.0
17.5
N (hadrons)

Toy example



"Experiment"
NF (base model)
NF (fine tuned)
Density
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
0.200
2.5
5.0
7.5
10.0
12.5
15.0
17.5
N (hadrons)

Toy example - comments

1. The solution provided by the network is not unique
    a. More observables should further constrain possible solutions

Toy example - comments

1. The solution provided by the network is not unique
    a. More observables should further constrain possible solutions
2. Because MAGIC relies on reweighting event samples, the target solution must have good coverage over the base model i.e. be a ‘perturbation’ of the base model

Conclusion

Invertible neural networks and the MAGIC training paradigm together allow for the training of microscopic first emission kinematics on macroscopic experimental distributions.

What's next: Full-fledged application ($e^+ e^-$)