Uncovering latent jet substructure

Barry M. Dillon
Jozef Stefan Institute, Ljubljana, Slovenia

Based on: **hep-ph/1904.04200**
BMD, Darius A. Faroughy, Jernej F. Kamenik

Dark Machines, Trieste, April 11th 2019

Institut
"Jožef Stefan"
Ljubljana, Slovenija

Overview

- Goal:
  Build an **unsupervised** ML tagger that can be used in **new physics** searches at colliders

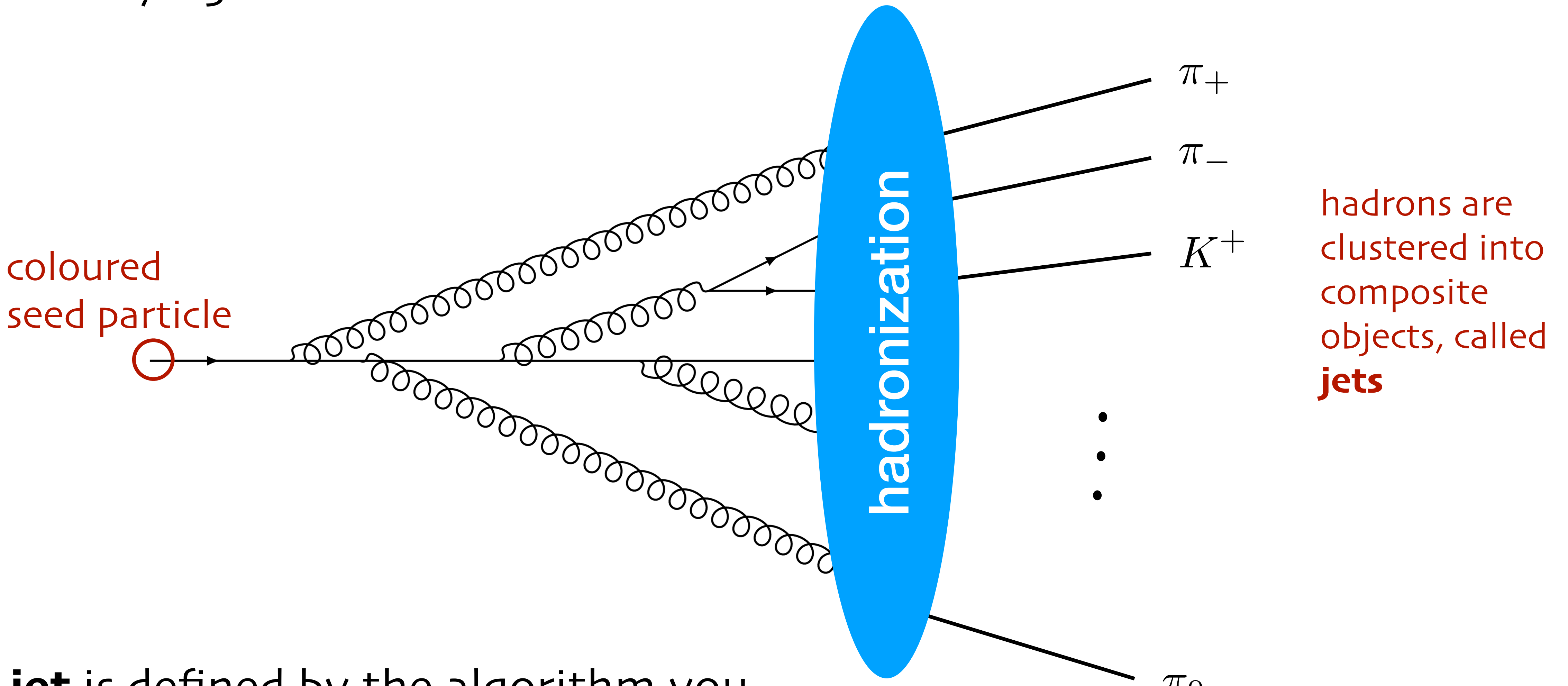
- How?
  Latent Dirichlet Allocation (LDA)

See talks:
'Probabilistic programming':
Rajat Mani Thomas
'Probabilistic Programming and Inference in Particle Physics':
Atılım Güneş Baydin

- Why?
  Model independence, data-driven, anomaly detection, you can see what the machine learned

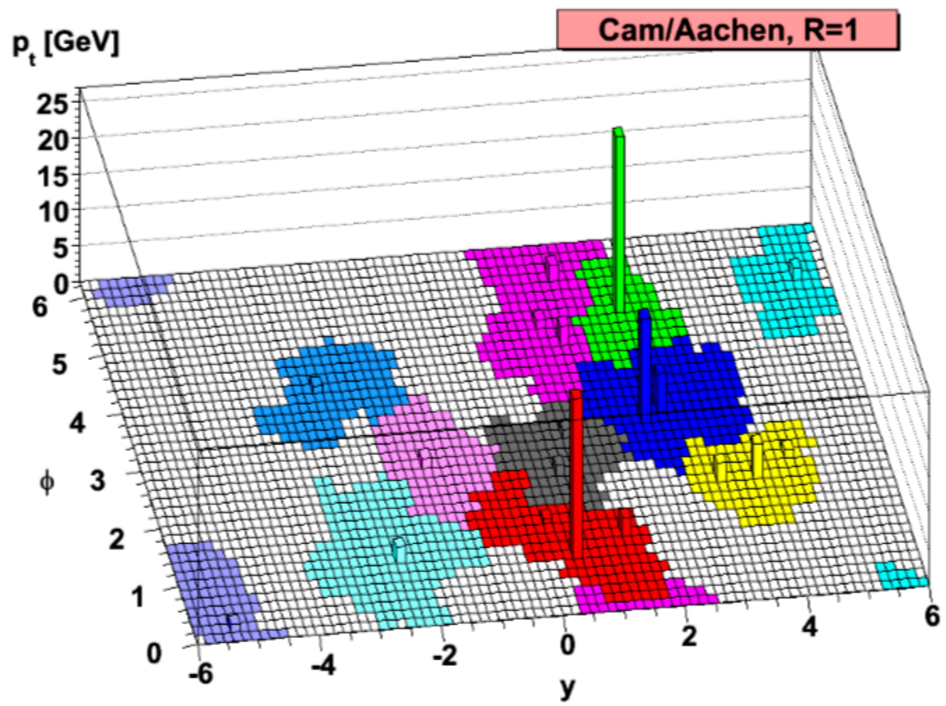
Jets and substructure

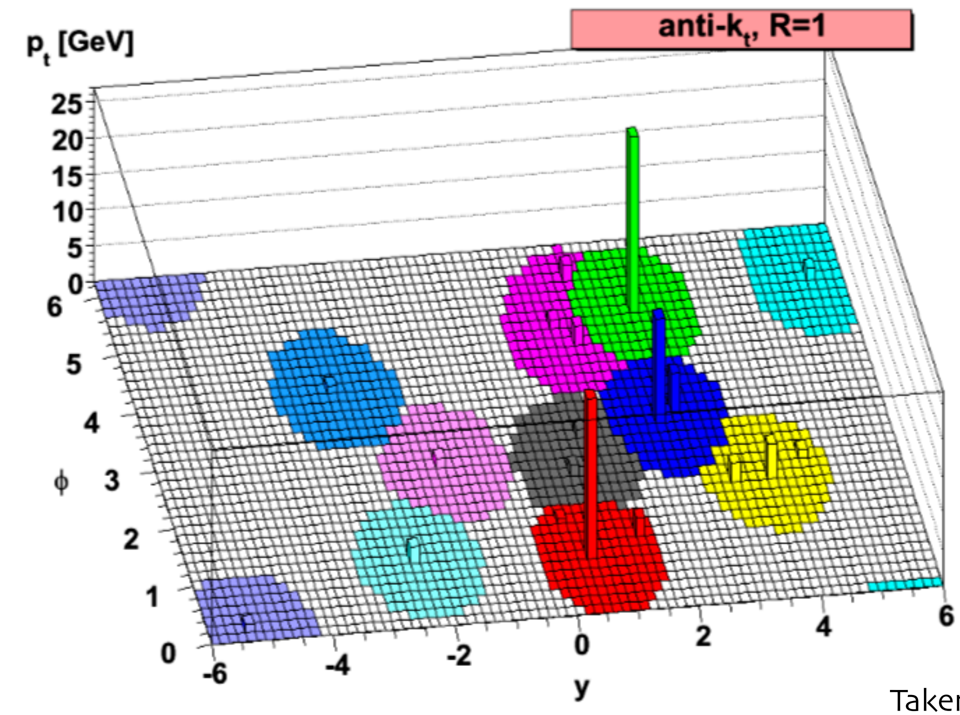
Events at colliders produce collimated bunch of hadrons initiated by some underlaying event:



a **jet** is defined by the algorithm you used to cluster the particles

Jets and substructure





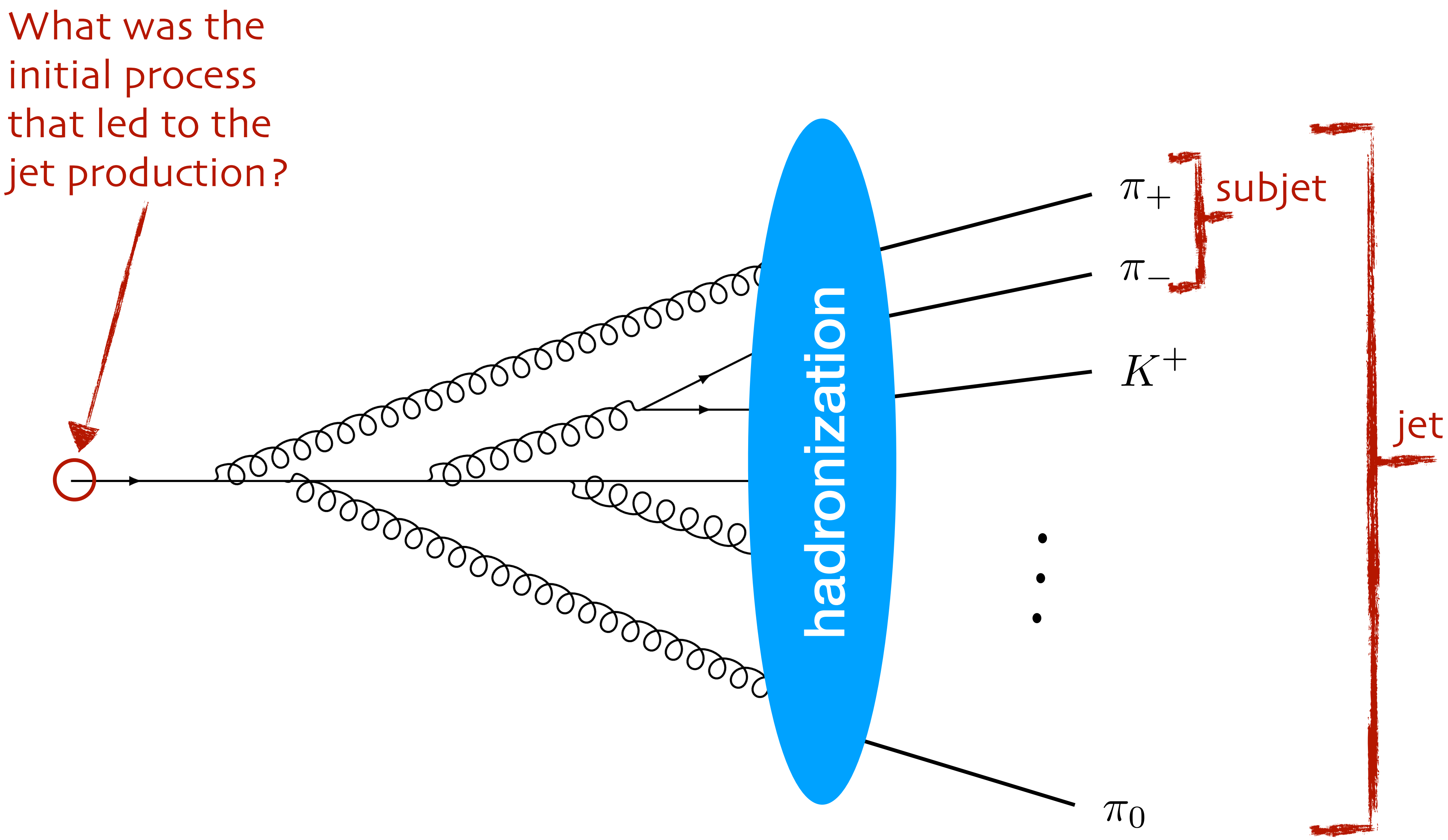
Taken from:
M. Cacciari, G. P. Salam, G. Soyez (2008)

Cambridge -Aachen

$$d_{ij} = \frac{\Delta R_{ij}^2}{R^2}, \quad d_{iB} = 1$$

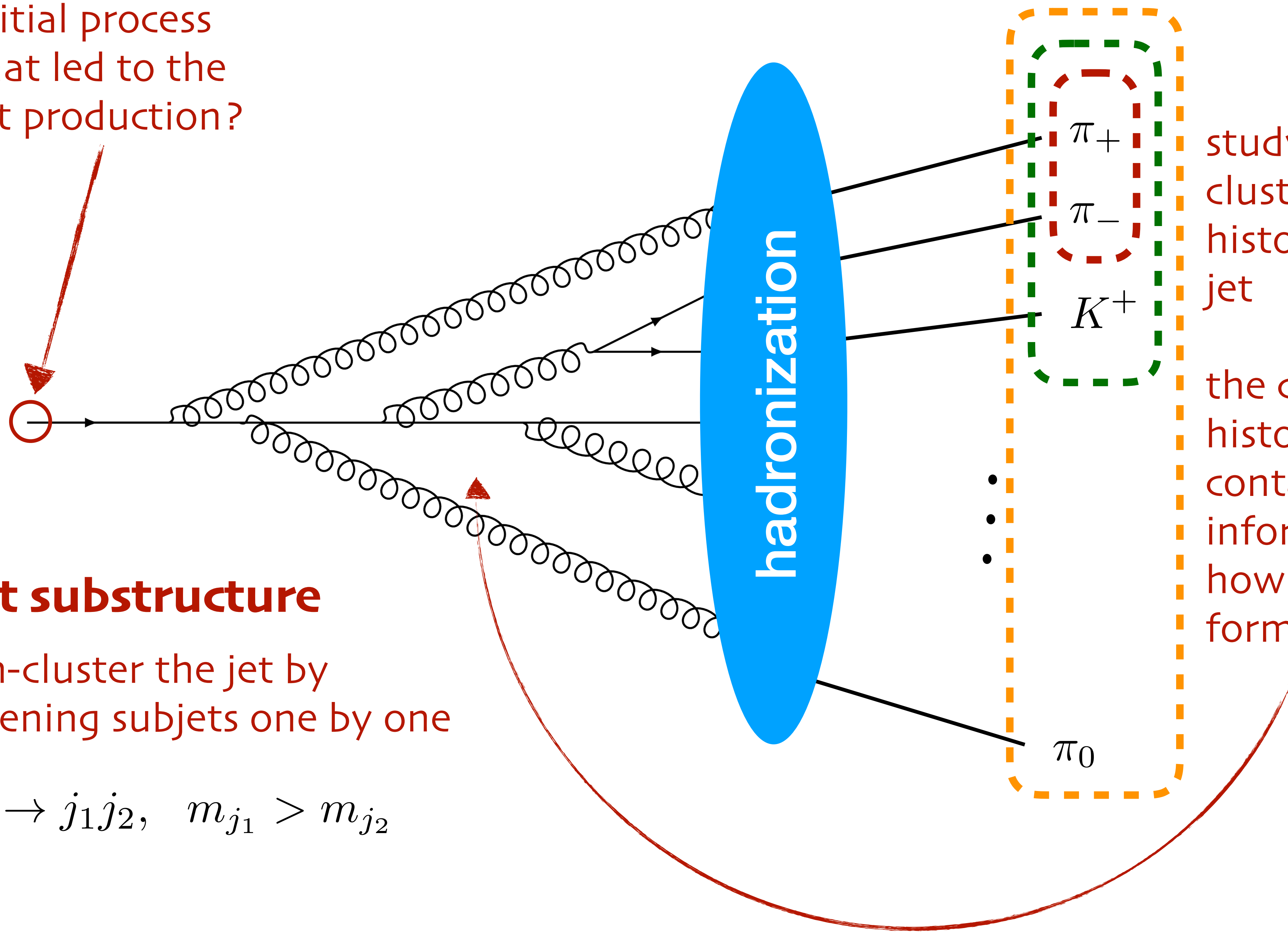
1 - compute d_{ij} for each particle in the final state
2 - if the minimum is d_{iB} declare particle i a jet, and remove it from the list
3 - if the minimum is d_{ij} combine particles i and j and go back to step 1
4 - repeat until there are no particles left

Jets and substructure



Jets and substructure

What was the initial process that led to the jet production?



study the clustering history of the jet

the clustering history contains information on how the jet formed

J. M. Butterworth, A. R. Davison, M. Rubin, G. P. Salam (2008)

Jet substructure

Un-cluster the jet by opening subjets one by one

$$j_0 \to j_1 j_2, \quad m_{j_1} > m_{j_2}$$

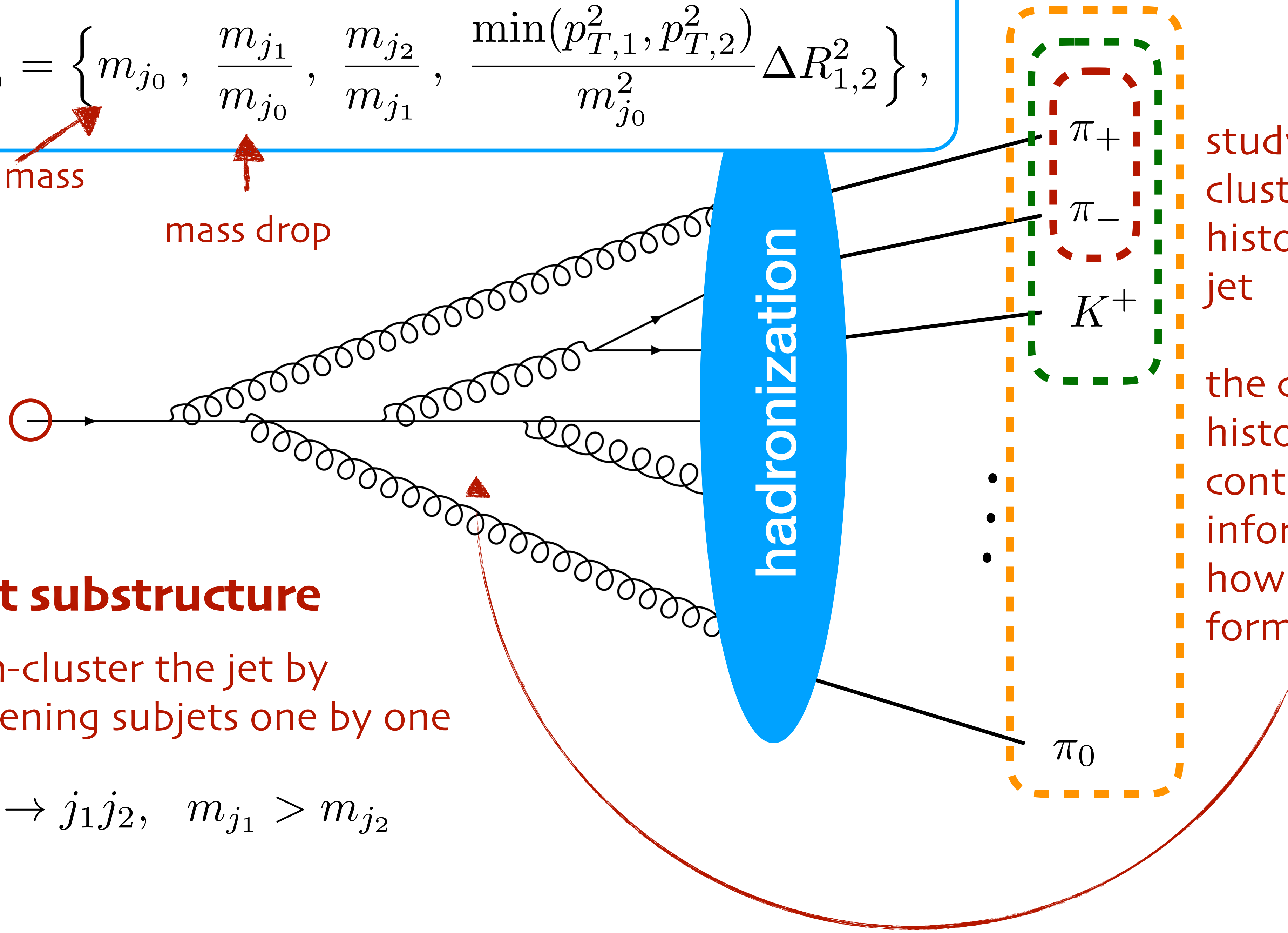
Jets and substructure

Useful substructure observables:

$$o_{j_0} = \left\{ m_{j_0} \, , \; \frac{m_{j_1}}{m_{j_0}} \, , \; \frac{m_{j_2}}{m_{j_1}} \, , \; \frac{\min(p_{T,1}^2, p_{T,2}^2)}{m_{j_0}^2} \Delta R_{1,2}^2 \right\} ,$$

subjet mass

mass drop



study the clustering history of the jet

the clustering history contains information on how the jet formed

J. M. Butterworth, A. R. Davison, M. Rubin, G. P. Salam (2008)

Jet substructure

Un-cluster the jet by opening subjets one by one

$$j_0 \to j_1 j_2 , \quad m_{j_1} > m_{j_2}$$

Top tagging

Top tagging: 'was this jet seeded by a top-quark or not?'

Signal: top jets from $t\bar{t}$ production in the SM

$$pp \to t\bar{t} \to jj, \quad (t \to W^+ b)$$

Features:
subjet mass
$m_{j_0} \sim m_t$ (175GeV)
$m_{j_0} \sim m_W$ (80GeV)
mass drop

$$\frac{m_{j_1}}{m_{j_0}} \sim \frac{m_W}{m_t} \sim 0.45$$

Background: QCD di-jets

$$pp \to gg \to jj$$

Features:
subjet mass
smoothly decaying distribution, peaked at zero
mass drop
smoothly decaying distribution, peaked at one

Tagging tops **manually** (e.g. the Johns-Hopkins (JH) top-tagger)

1 - cluster with C/A and then uncluster
2 - cuts are applied manually to filter out jets which have top-like features

D. E. Kaplan, K. Rehermann, M. D. Schwartz and B. Tweedie (2008)

Latent Dirichlet Allocation

D. M. Blei, A. Y. Ng, M. I. Jordan, J. Lafferty (2003)

Characterising documents as a set of 'topics' or 'themes'

LDA is based on a generative process for writing documents

Assumptions: short distance physics is represented by a set of 'themes'

A 'theme' is a distribution over substructure features

a jet, or event, is represented by a list (document) of features

each jet, or event, can have different proportions of each theme

A mixed sample of jets or events can be parameterised by a set of 'latent' hyper-parameters:

α_i theme concentration parameters

β_{ij} theme-feature matrix

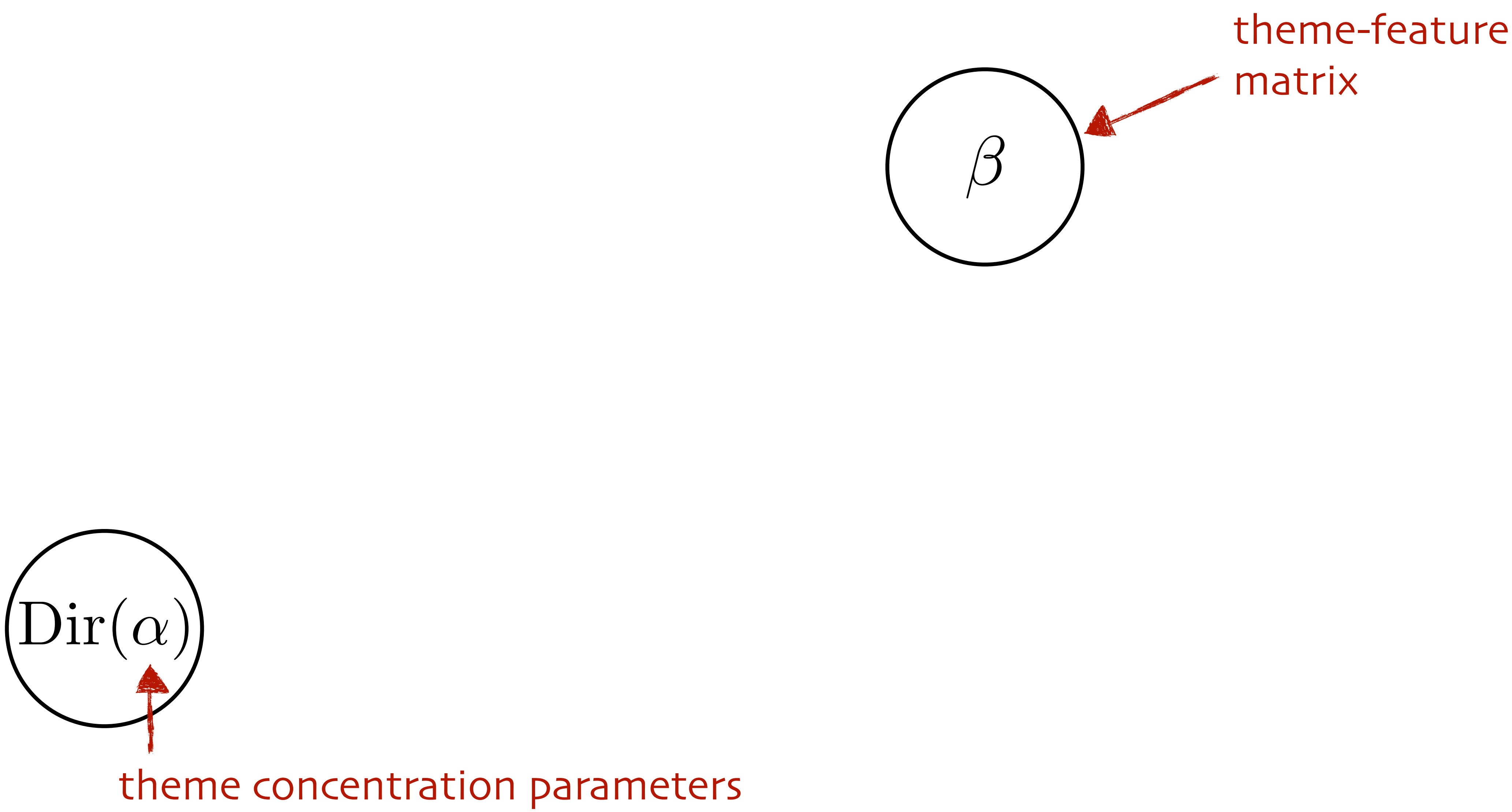
$i = 1, \ldots, K$ ← #themes (finite)

$j = 1, \ldots, N_f$ ← #features

Latent Dirichlet Allocation

D. M. Blei, A. Y. Ng, M. I. Jordan, J. Lafferty (2003)

The LDA process for generating jets or events:



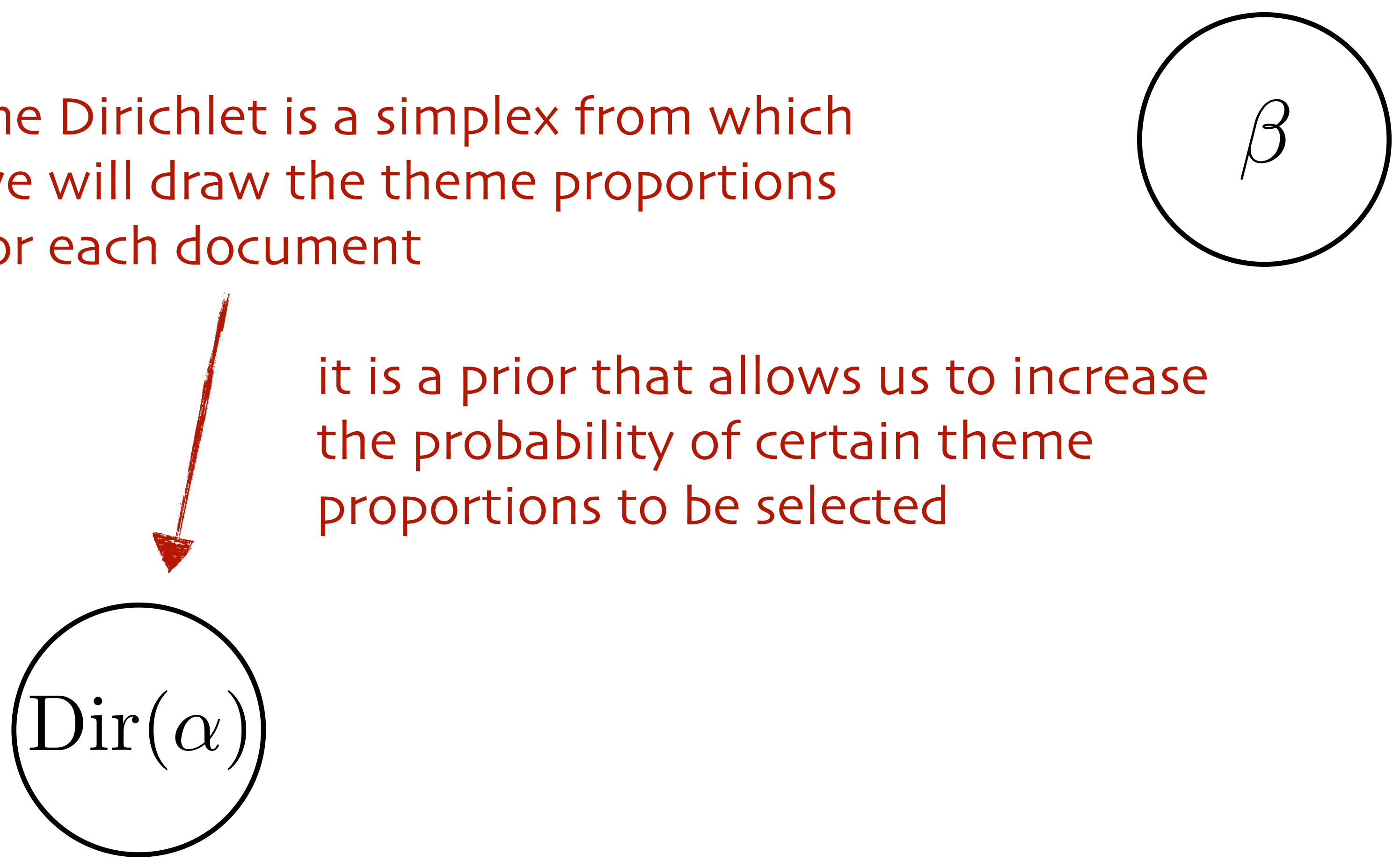
Latent Dirichlet Allocation

D. M. Blei, A. Y. Ng, M. I. Jordan, J. Lafferty (2003)

The LDA process for generating jets or events:

the Dirichlet is a simplex from which we will draw the theme proportions for each document

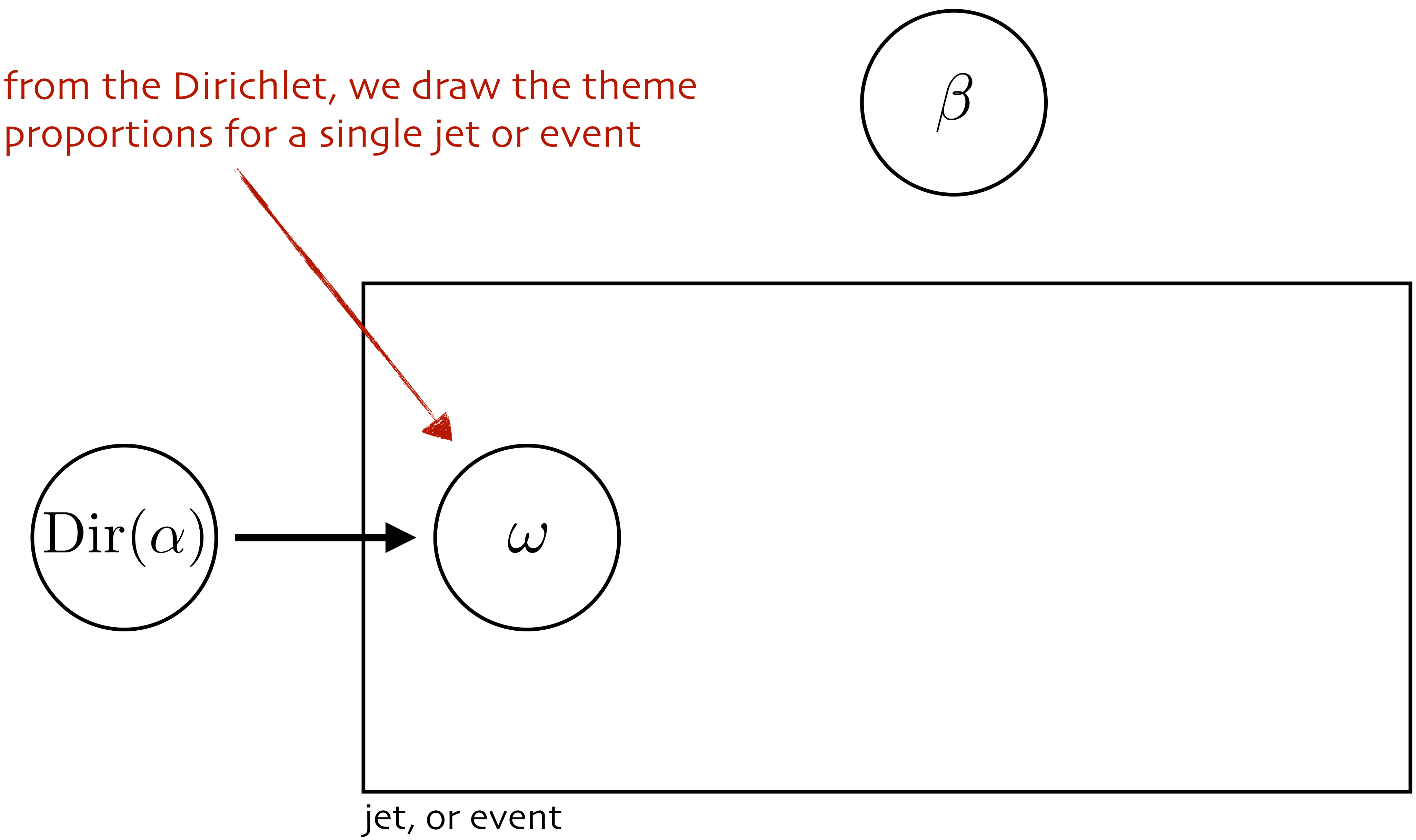
it is a prior that allows us to increase the probability of certain theme proportions to be selected



Latent Dirichlet Allocation

D. M. Blei, A. Y. Ng, M. I. Jordan, J. Lafferty (2003)

The LDA process for generating jets or events:

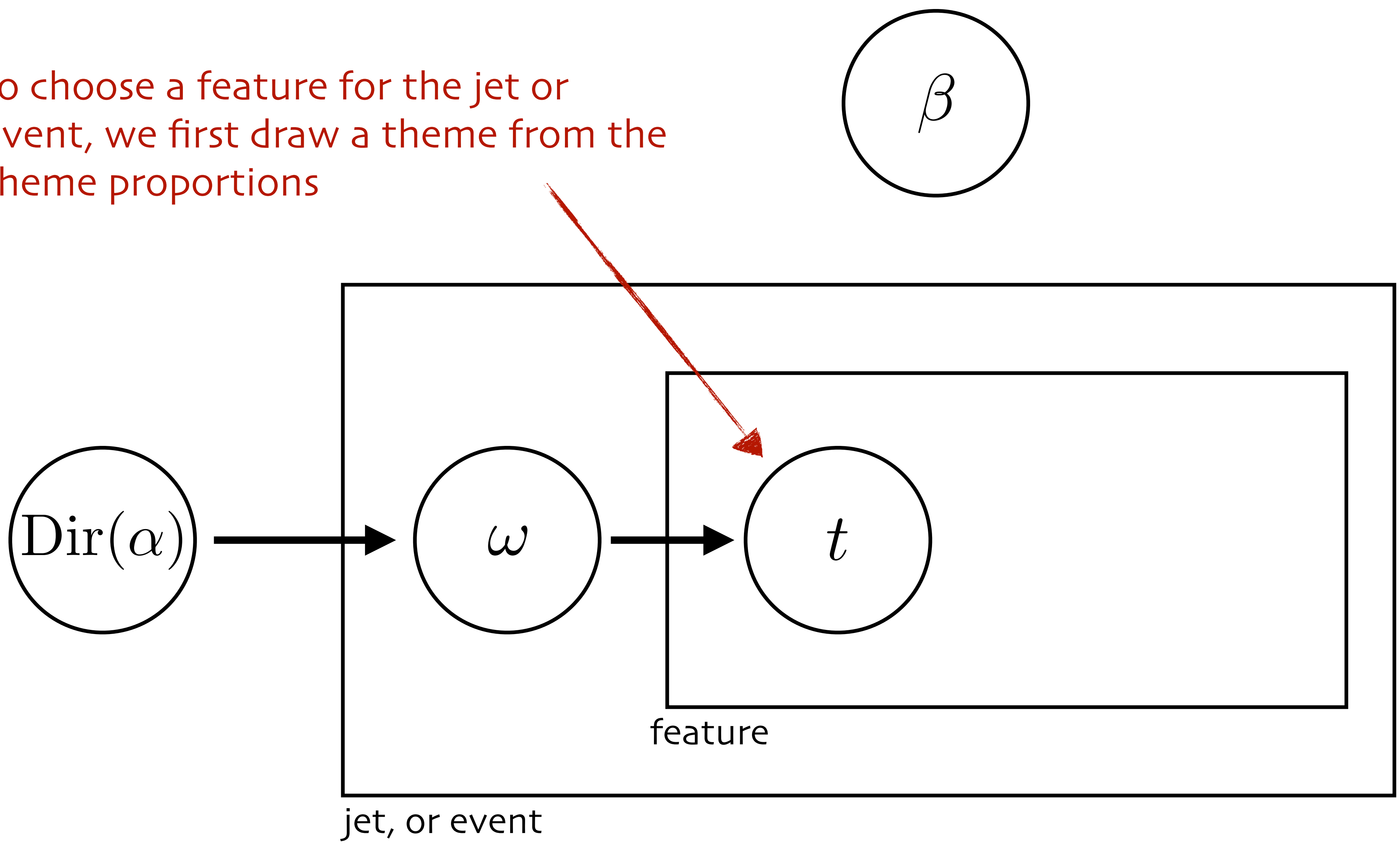


Latent Dirichlet Allocation

D. M. Blei, A. Y. Ng, M. I. Jordan, J. Lafferty (2003)

The LDA process for generating jets or events:

to choose a feature for the jet or event, we first draw a theme from the theme proportions

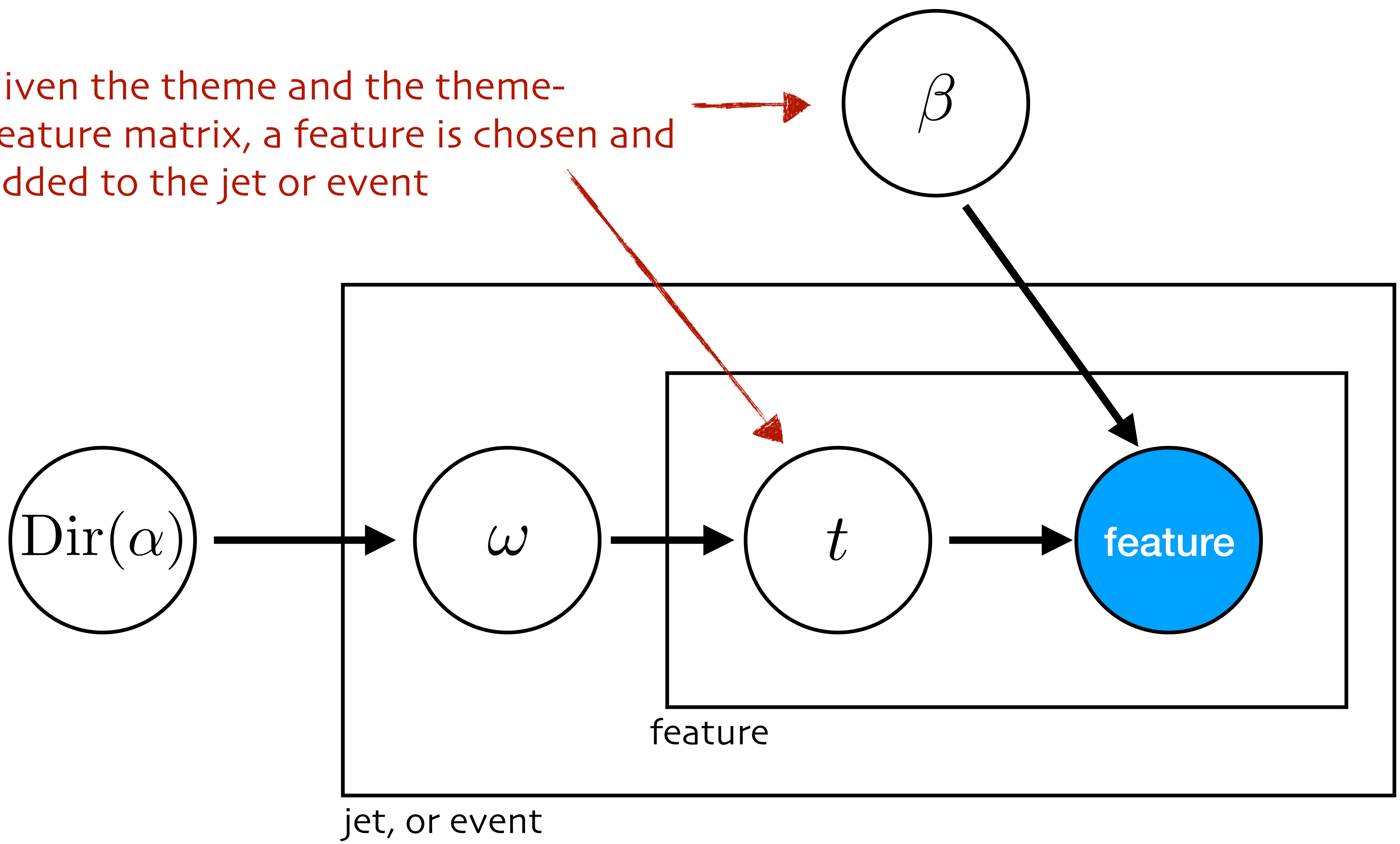


Latent Dirichlet Allocation

D. M. Blei, A. Y. Ng, M. I. Jordan, J. Lafferty (2003)

The LDA process for generating jets or events:

given the theme and the theme-feature matrix, a feature is chosen and added to the jet or event

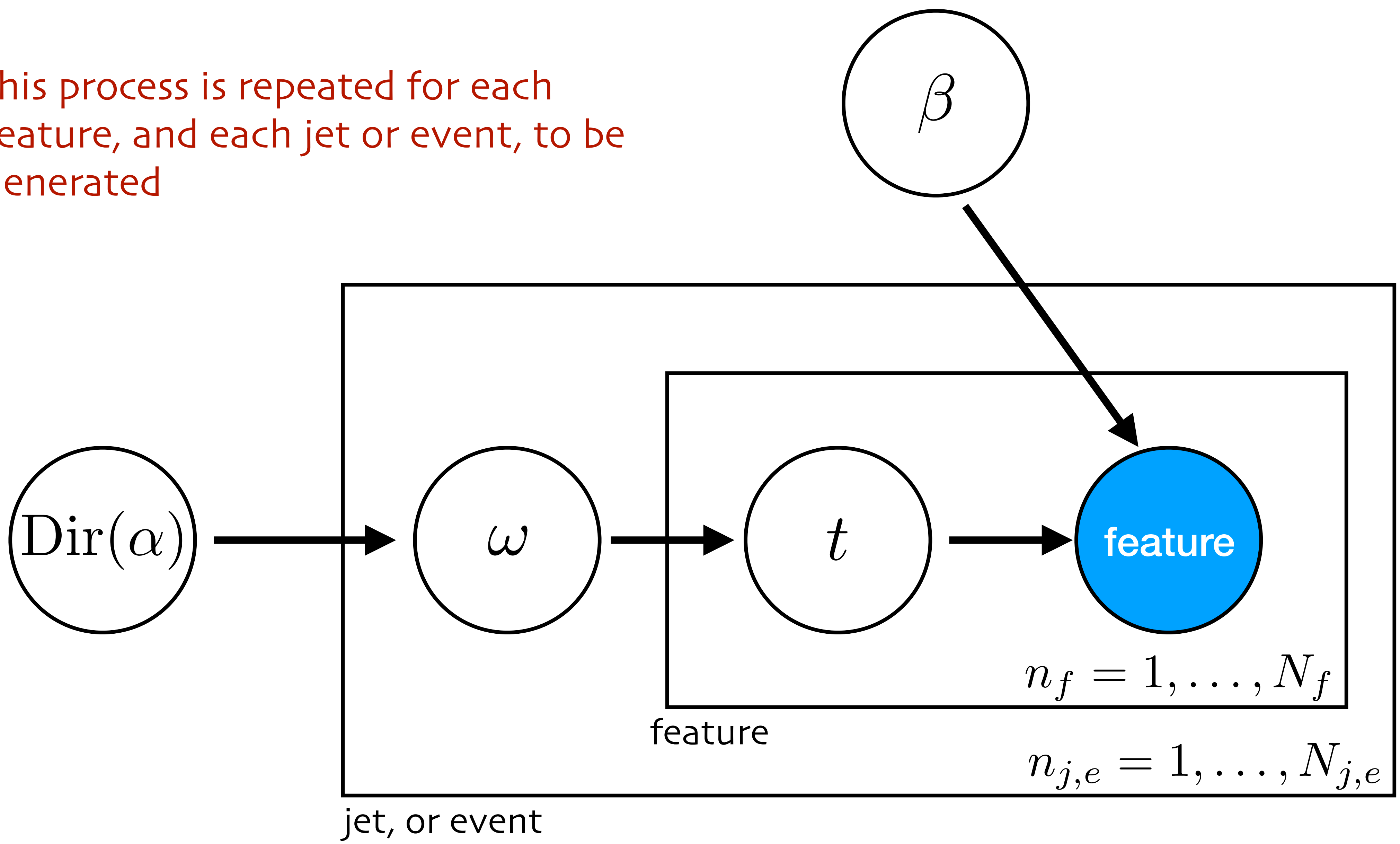


Latent Dirichlet Allocation

D. M. Blei, A. Y. Ng, M. I. Jordan, J. Lafferty (2003)

The LDA process for generating jets or events:

this process is repeated for each feature, and each jet or event, to be generated



Latent Dirichlet Allocation

D. M. Blei, A. Y. Ng, M. I. Jordan, J. Lafferty (2003)

The probability of a jet being generated, given the choice of latent parameters, is

$$p(j|\alpha,\beta) = \int_\omega p(\omega|\alpha) \prod_{f\in j} \left(\sum_t p(t|\omega) p(f|t,\beta) \right)$$

The goal: to **infer the latent parameters** in the theme-feature matrix, by analysing a collection of documents

How? Variational Bayesian methods, implemented using the gensim software

R. Rehurek, P. Sojka (2010)
M. D. Hoffman, D. M. Blei, F. Bach (2010)

Latent Dirichlet Allocation

D. M. Blei, A. Y. Ng, M. I. Jordan, J. Lafferty (2003)

The probability of a jet being generated, given the choice of latent parameters, is

$$p(j|\alpha,\beta) = \int_\omega p(\omega|\alpha) \prod_{f\in j} \left(\sum_t p(t|\omega) p(f|t,\beta) \right)$$

The goal: to **infer the latent parameters** in the theme-feature matrix, by analysing a collection of documents

How? Variational Bayesian methods, implemented using the gensim software

R. Rehurek, P. Sojka (2010)
M. D. Hoffman, D. M. Blei, F. Bach (2010)

Given a collection of jets or events, we can choose a number of themes, and α_i, then the LDA algorithm estimates the latent β_{ij} .
We can disentangle short distance physics based on their features in the jet substructure, in an unsupervised way.

Latent Dirichlet Allocation

D. M. Blei, A. Y. Ng, M. I. Jordan, J. Lafferty (2003)

Useful substructure observables:

$$o_{j_0} = \left\{ m_{j_0} \, , \; \frac{m_{j_1}}{m_{j_0}} \, , \; \frac{m_{j_2}}{m_{j_1}} \, , \; \frac{\min(p_{T,1}^2, p_{T,2}^2)}{m_{j_0}^2} \Delta R_{1,2}^2 \right\} ,$$

this is a feature in the substructure

1 - un-cluster the jet, calculate the above observables at each stage

2 - bin the observables, and form a feature for each stage, from the observables

3 - form a 'document' describing each jet, and a mixed sample of different jets

4 - analyse these documents using LDA - find the 'themes' describing the physics

5 - use inference to identify themes in new jets - identify the origin of the jet

LDA top tagging

For our study:

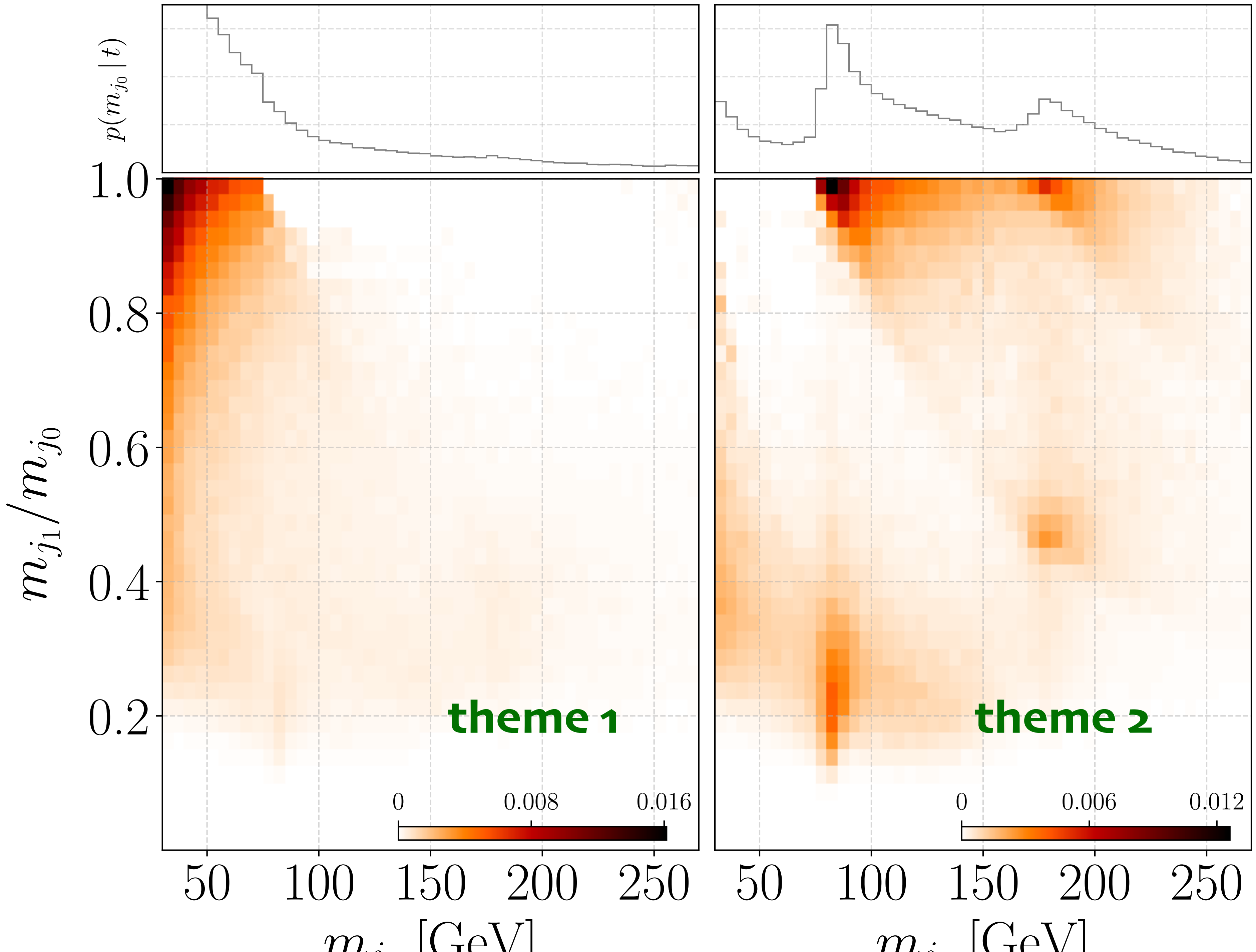
1 - train LDA on mixed samples: $S/B = 1,\ 1/9,\ 1/99$

2 - $p_T \in [350, 450]$ GeV

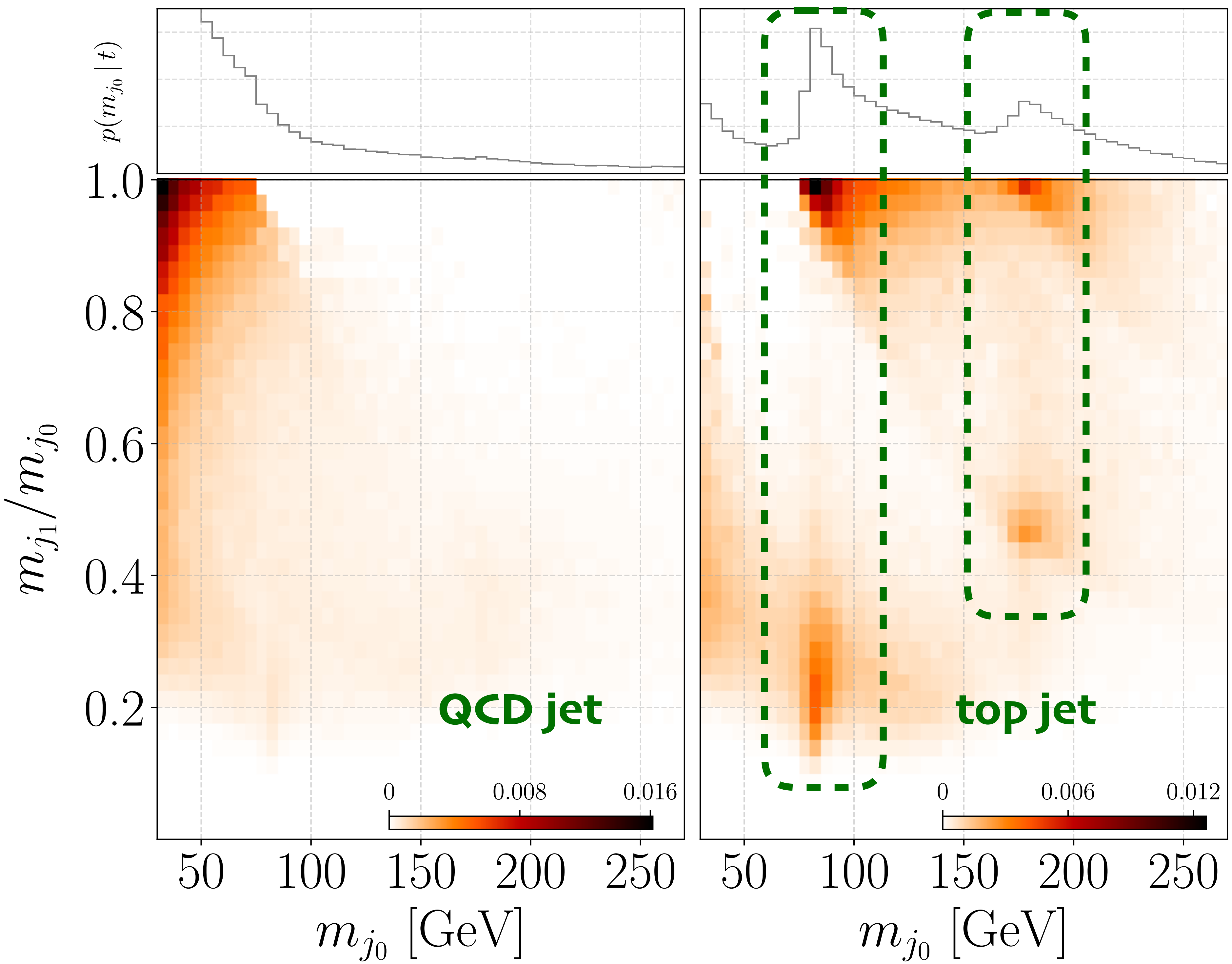
3 - sample size: $\sim 8 \times 10^4$

4 - in accordance with S/B: $\alpha = [0.5, 0.5],\ [0.9, 0.1],\ [0.99, 0, 01]$

LDA top tagging

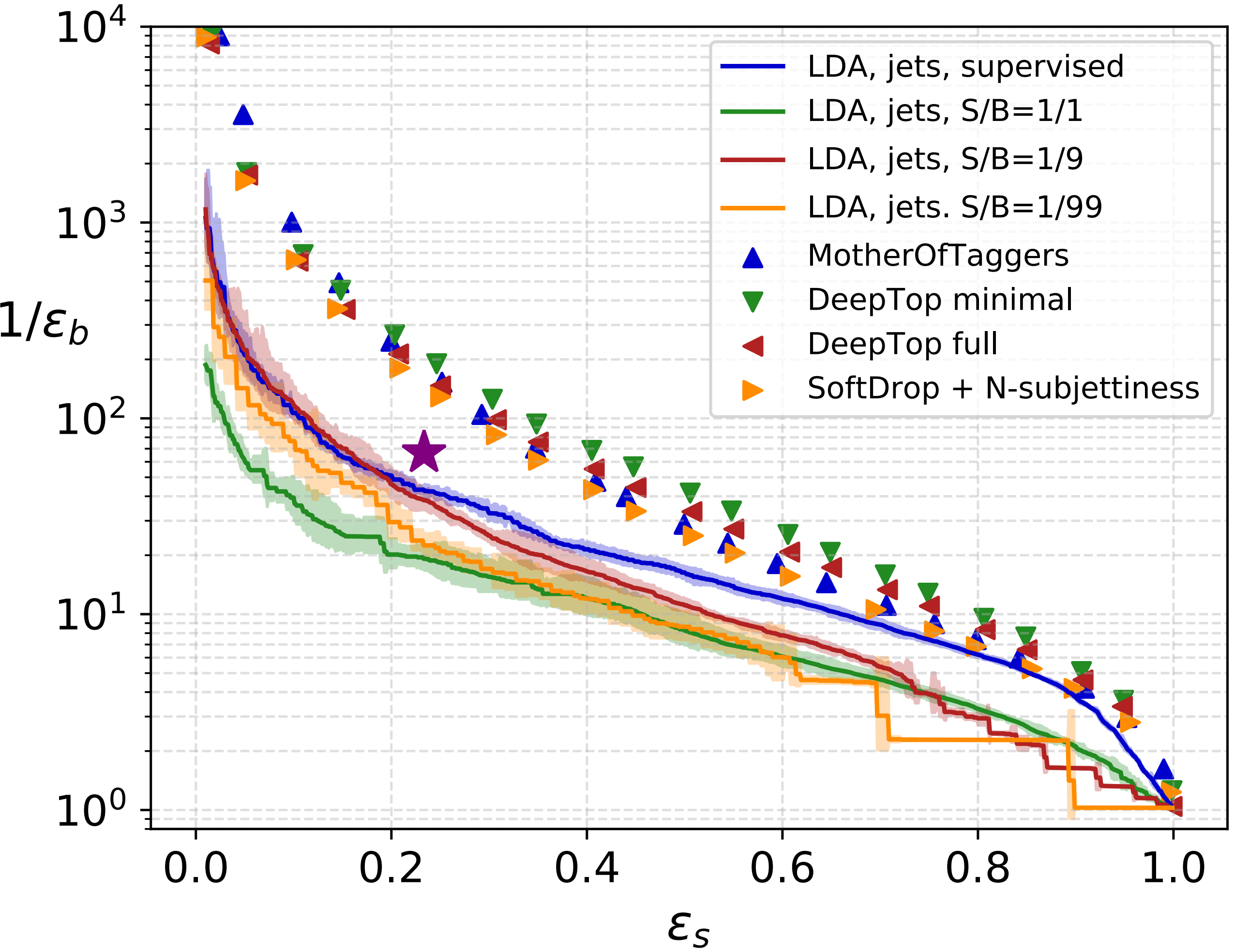


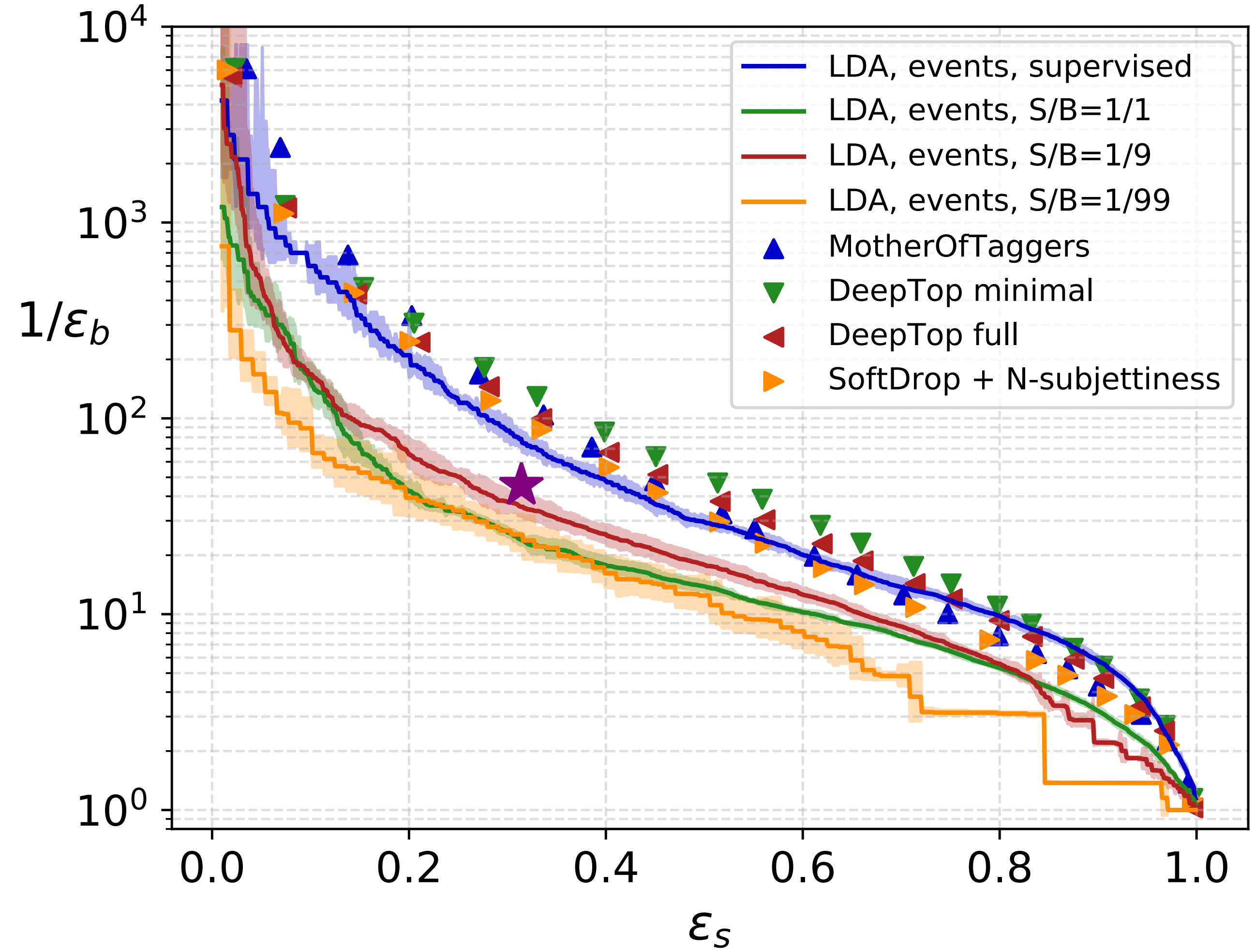
LDA top tagging



LDA top tagging

Measure performance with ROC curves:





G. Kasieczka, T. Plehn, M. Russell, T. Schell (2017)

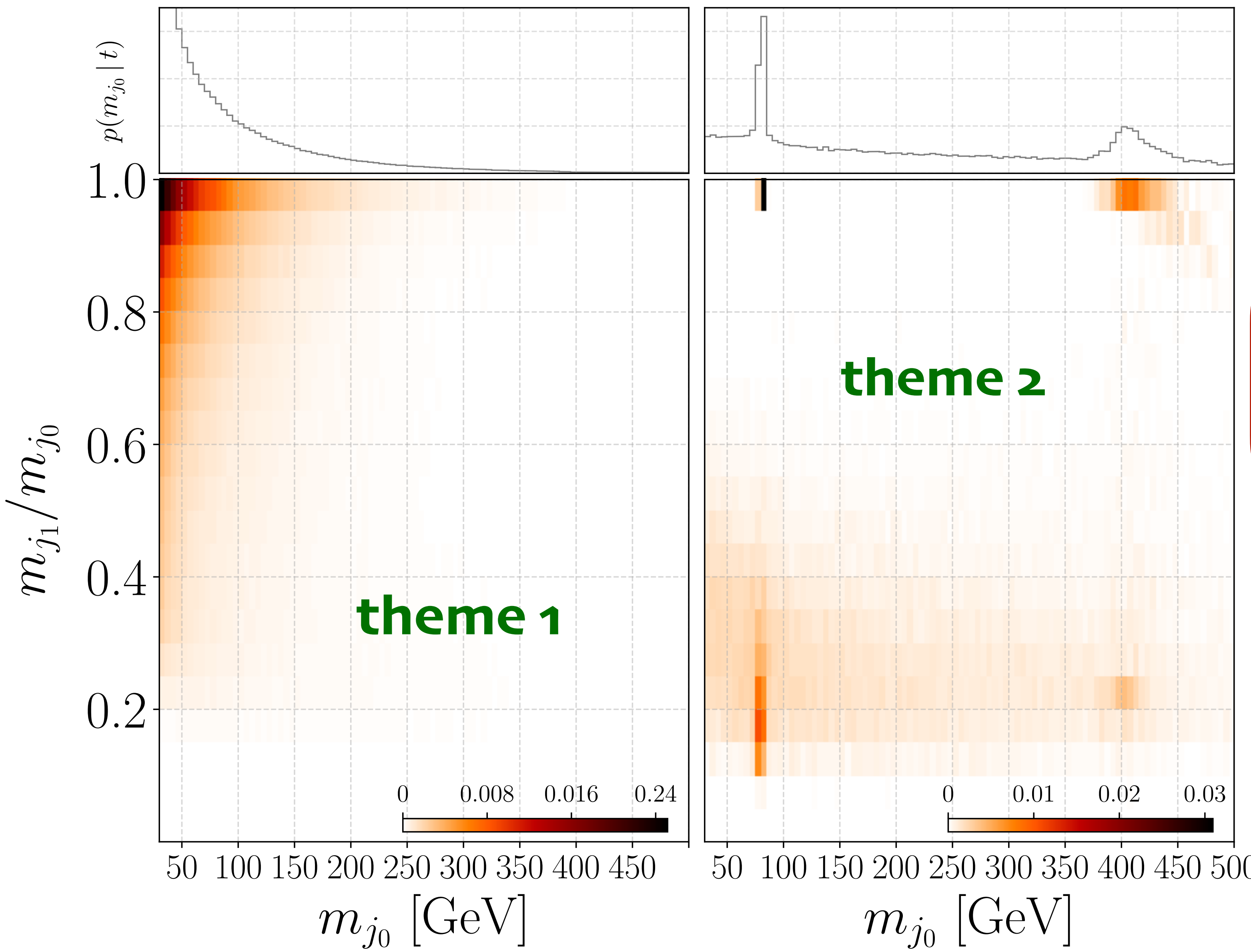
results compared to JH top tagger (purple star) and DeepTop

results have been k-folded, k=10, to estimate robustness

LDA **new physics** tagging

Now for a NP process: $pp \to W' \to \phi W \to WWW$

$m_{W'} = 3 \text{ TeV}, \; m_\phi = 400 \text{ GeV}$



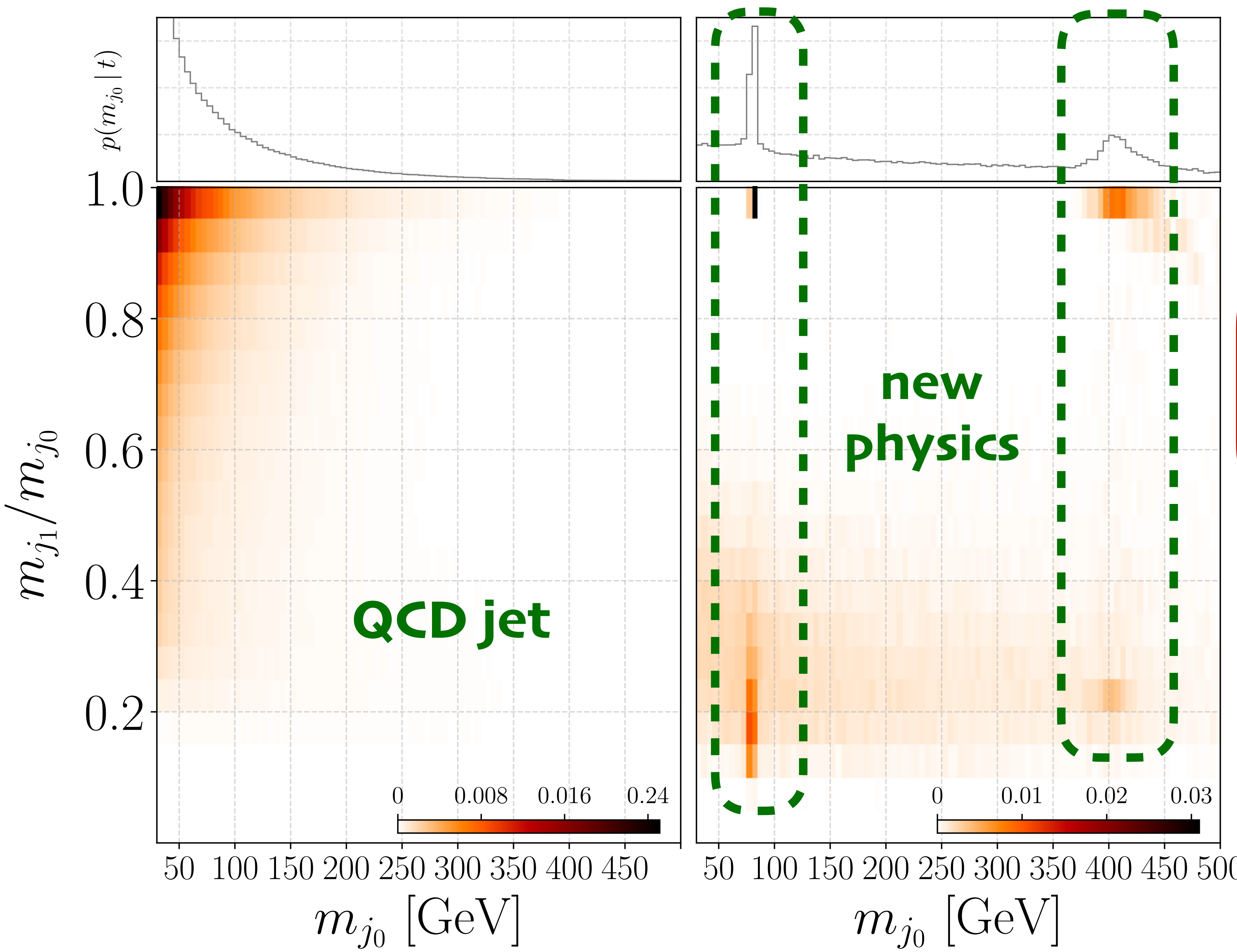
$S/B = 0.011$

$\alpha = [0.989, 0.011]$

LDA **new physics** tagging

Now for a NP process: $pp \to W' \to \phi W \to WWW$

$m_{W'} = 3 \text{ TeV}, \; m_\phi = 400 \text{ GeV}$

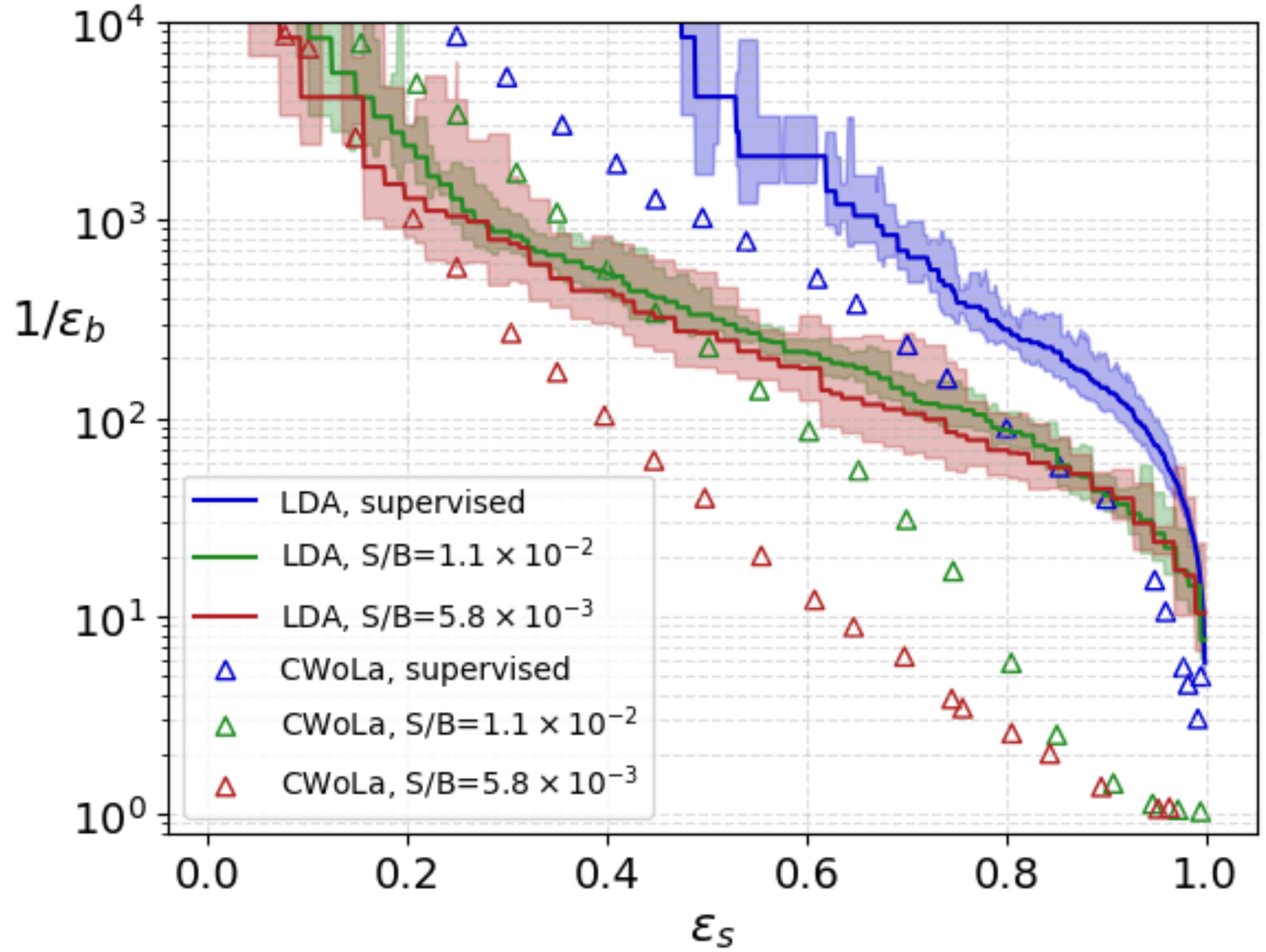


$S/B = 0.011$

$\alpha = [0.989, 0.011]$

LDA **new physics** tagging

Measure performance with ROC curves:



results compared to CWoLa tagger

J. H. Collins, K. Howe, B. Nachman (2019)

results have been k-folded, k=10, to estimate robustness

Summary and next steps

- We use LDA as an unsupervised algorithm for disentangling signal and background events even at low S/B

- The algorithm characterises physical features associated to S and B, we can see what the algorithm learns

- The one algorithm can be used as a multi-purpose tagger:
  tops, W', other new physics

- Next steps:
  - use more observables in tagging (n-subjettiness, jet shapes, …)
  - find a way to fix hyper-parameters without knowing S/B
  - implement an LDA anomaly detector
  - expand beyond di-jets, to signals interesting for DM
  - use this algorithm in an unsupervised new physics search