Uncovering latent jet substructure

Jernej F. Kamenik
mostly based on 1904.04200
with D. A. Faroughy & B. M. Dillon + Manuel Szewc

Institut "Jožef Stefan" Ljubljana, Slovenija

Univerza v Ljubljani
Fakulteta za matematiko in fiziko

HC2NP
Tenerife, 23-28 September 2019

The Challenge

How to disentangle NP from NP?

The Challenge

How to disentangle NP from NP?

QCD UV

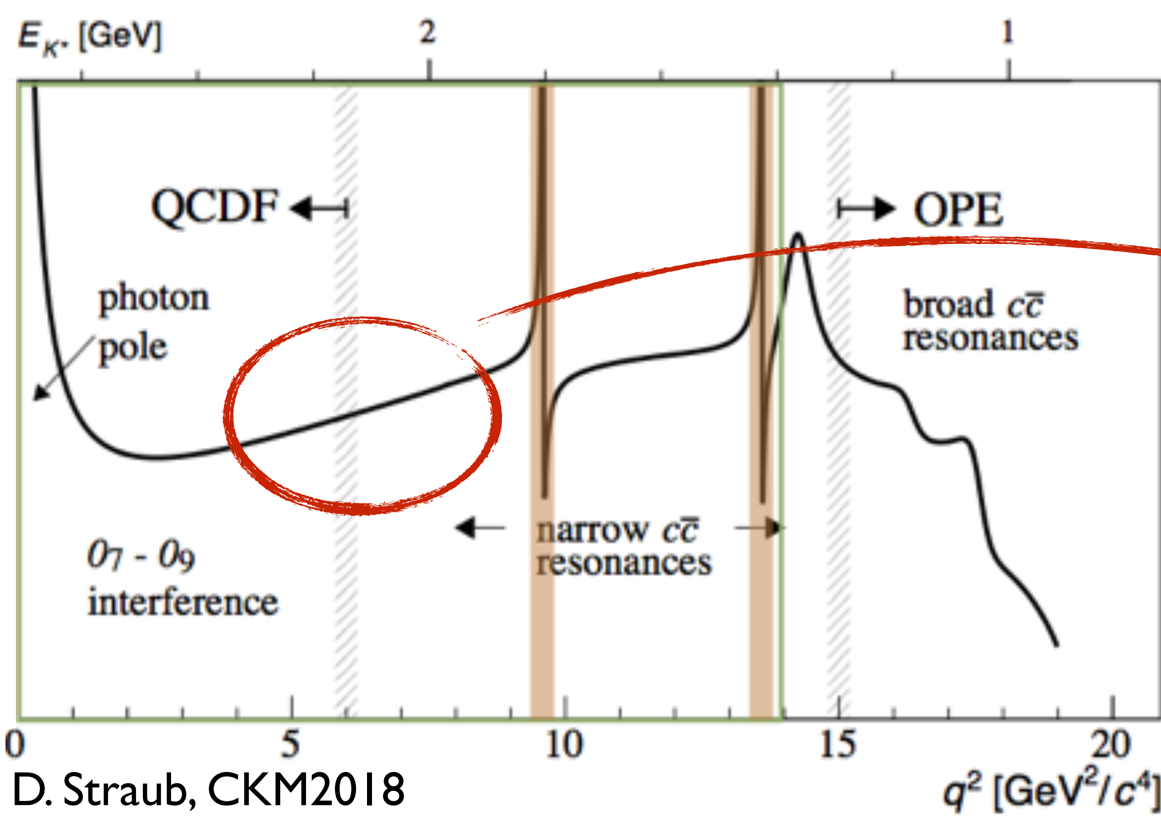
The Challenge

How to disentangle NP from NP?

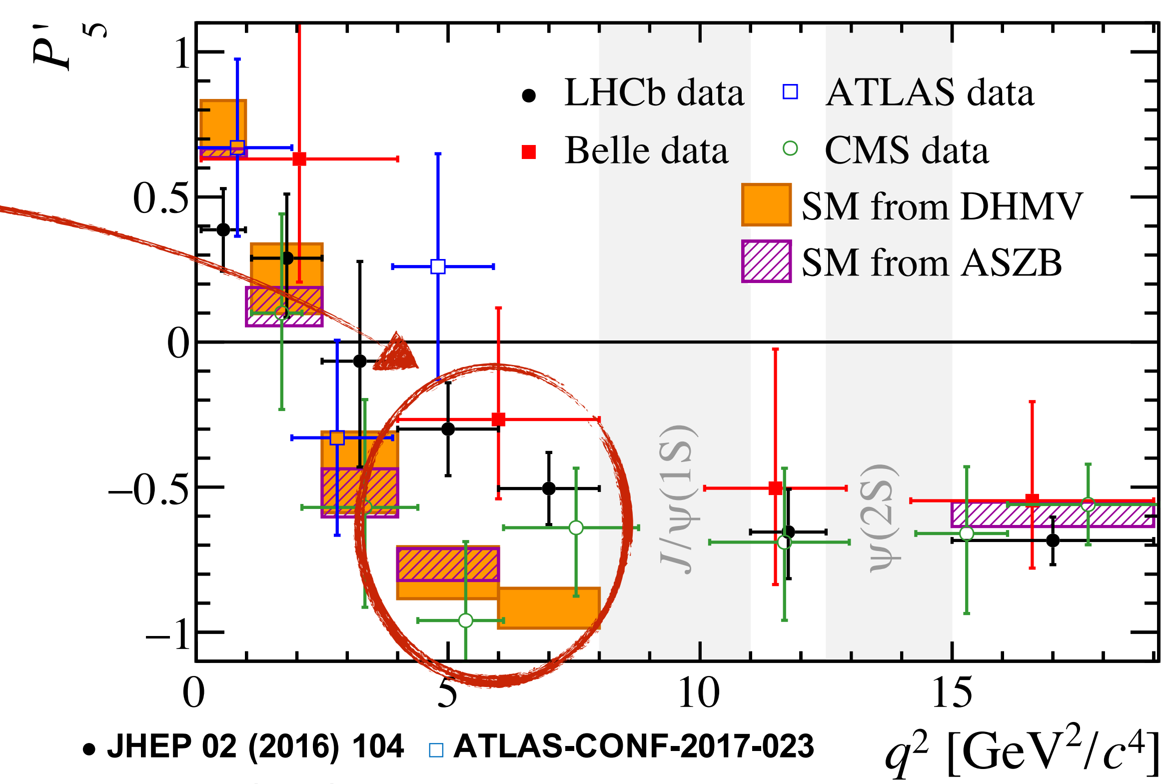
QCD UV

<u>Example</u>: spectrum of $B \to K^* \ell^+ \ell^-$

also $(g-2)_\mu$, ε'/ε, ...



D. Straub, CKM2018



The Challenge

How to disentangle NP from NP?

QCD UV

Example: origin of jets at LHC

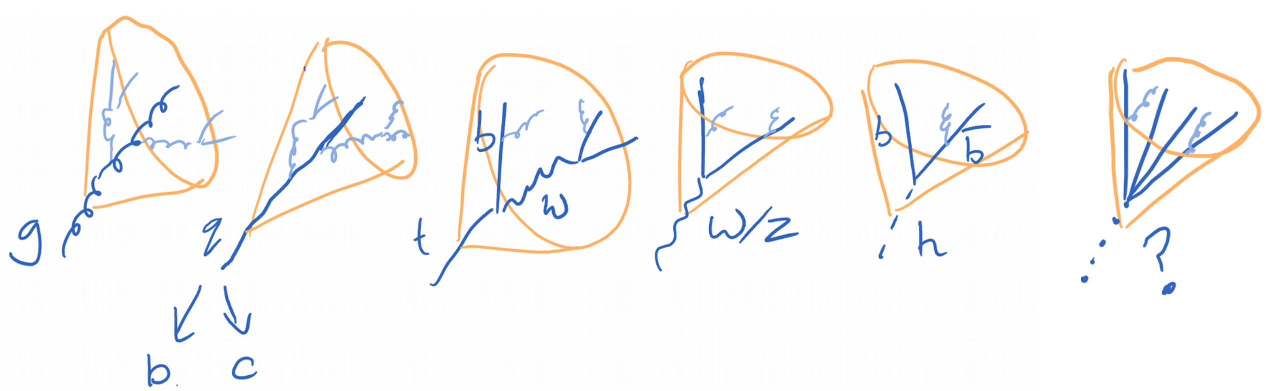
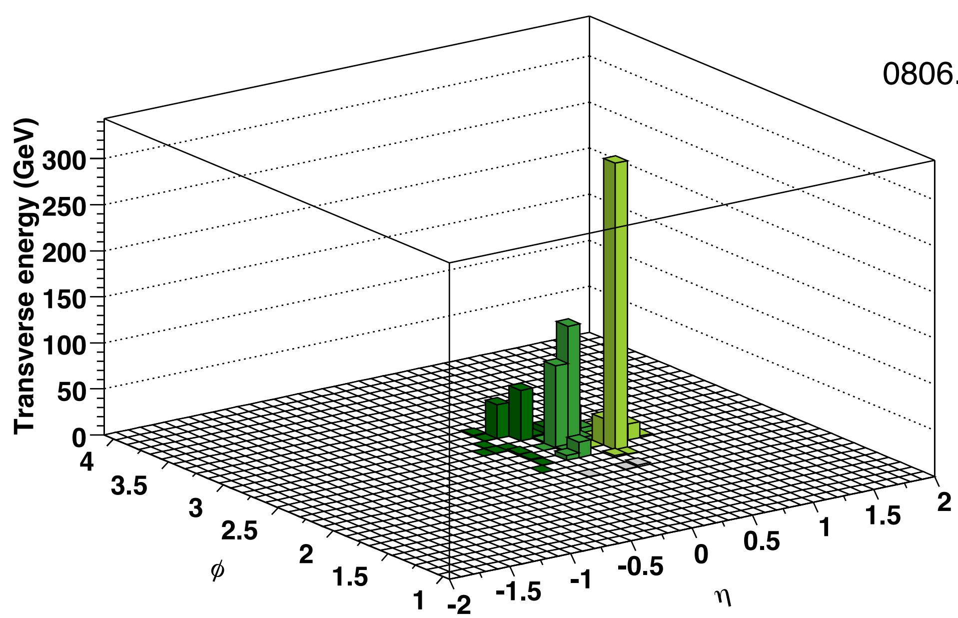


figure by J. Collins



0806.0848

The Challenge

How to disentangle NP from NP?

QCD UV

Example: origin of jets at LHC

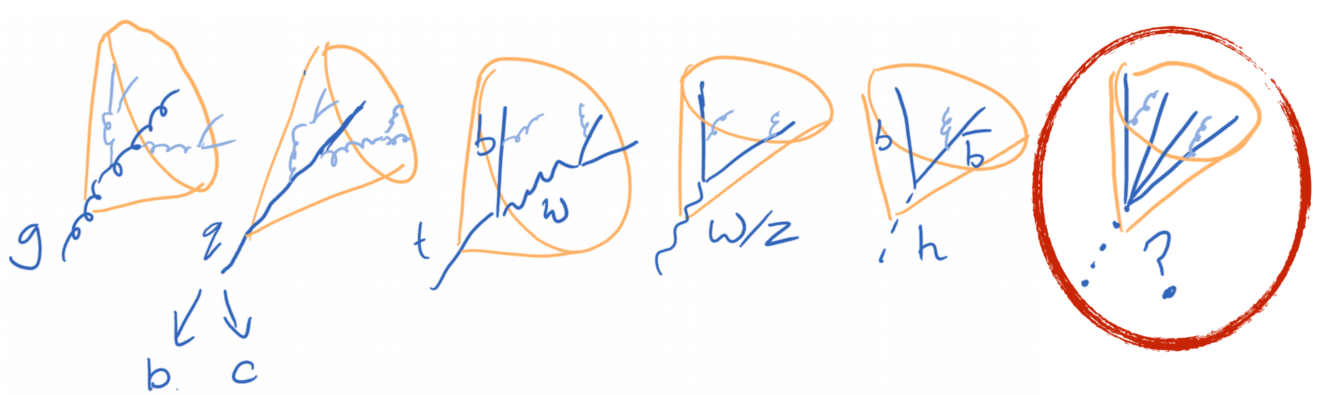
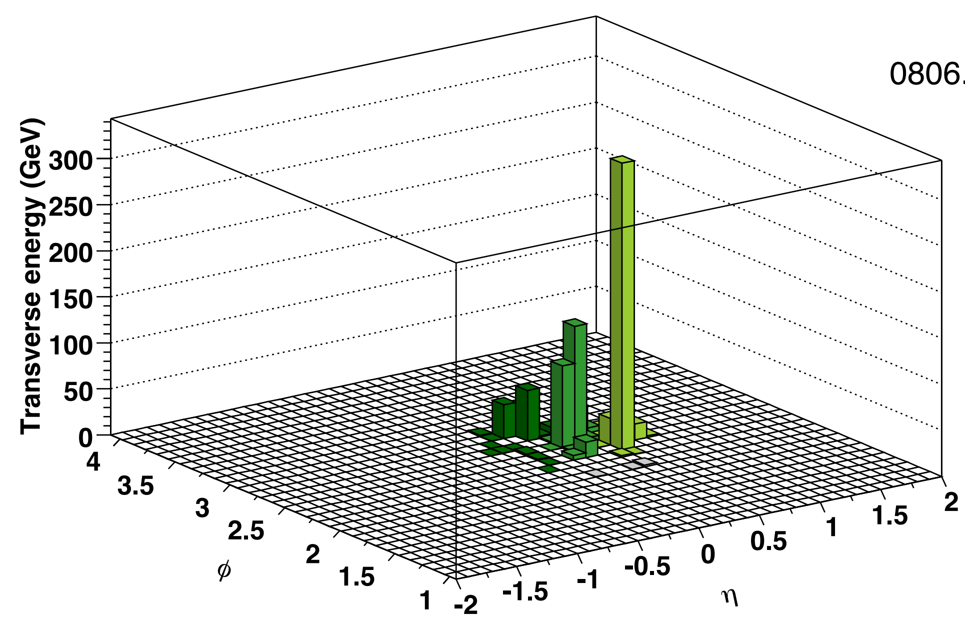


figure by J. Collins



0806.0848

The Challenge

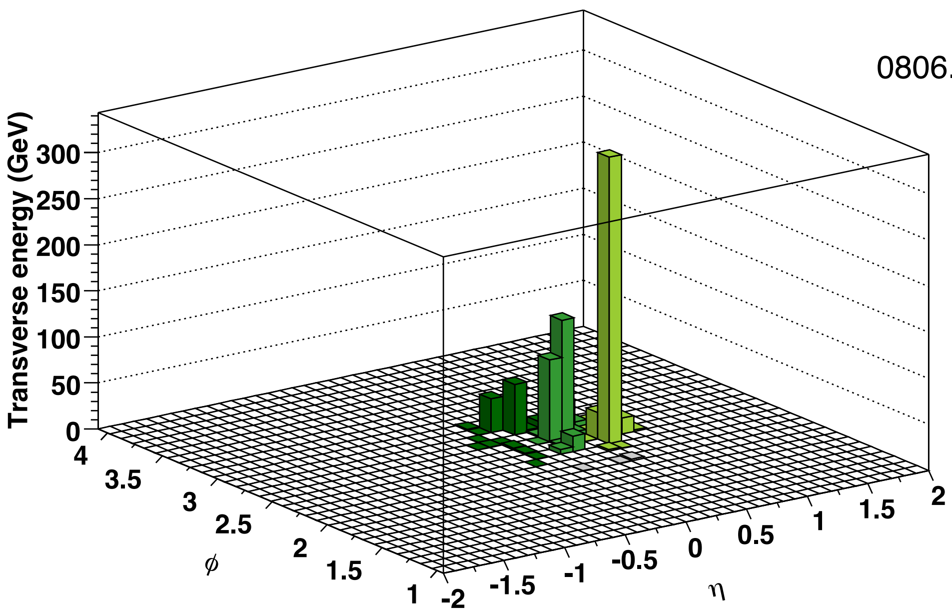
How to disentangle NP from NP?

QCD UV

<u>Example</u>: origin of jets at LHC

Generic NP cascades:

$H^+ \to t\,b$ (hadronic + leptonic)
$X \to YY \to hhhh$
$X \to WW$ (hadronic + semi-leptonic)
$X \to AB \to \tau\tau gg$
....



O(300) possibilities for boosted $X \to 4$ SM particles O(10k) possibilities for dijet combinations.

Dark quarks with BSM showers?

see e.g. Schwaller, Stolarski & Weiler, 1502.05409

The Challenge

How to disentangle NP from NP?

QCD UV

Task of classification

QCD Limited theoretical control & tools (pert. QCD, QCDF, sum rules, lattice, QCD shower & fragmentation models, …)

UV Exhaustive exploration of NP model space?

+ experimental systematics, limited statistics, etc…

How much can "we" learn directly from data?

Approaches to classification

1. Supervised (boosted decision trees, neural networks, …)

   train (general enough) model on (fit to) pre-labelled data

   domain (QCD) knowledge through labelling/generation of training data

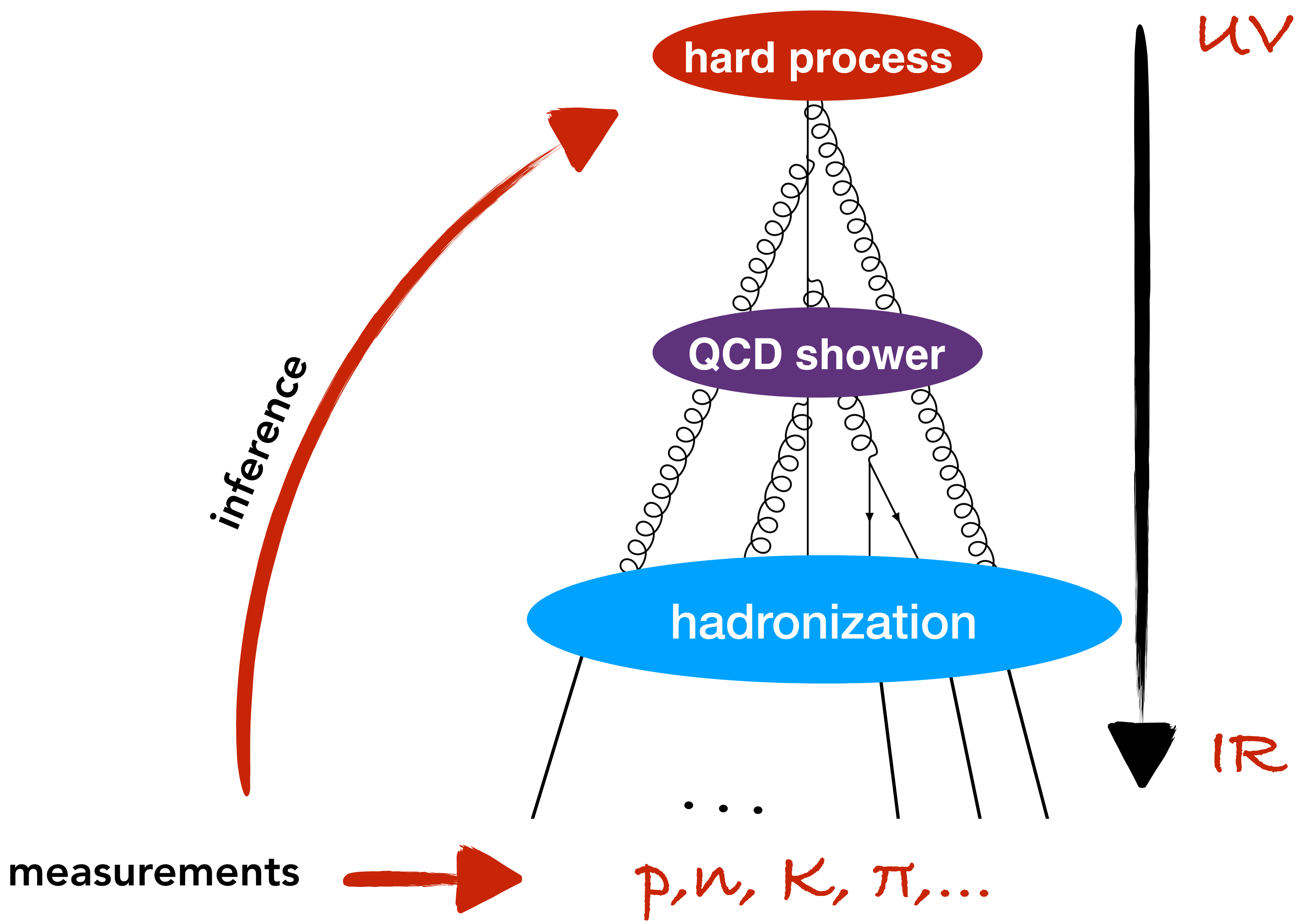
2. Unsupervised (e.g. (jet) clustering algorithms)

   search for structures within (unlabelled) data using priors for the structure distributions

   domain knowledge through choice of priors / clustering model

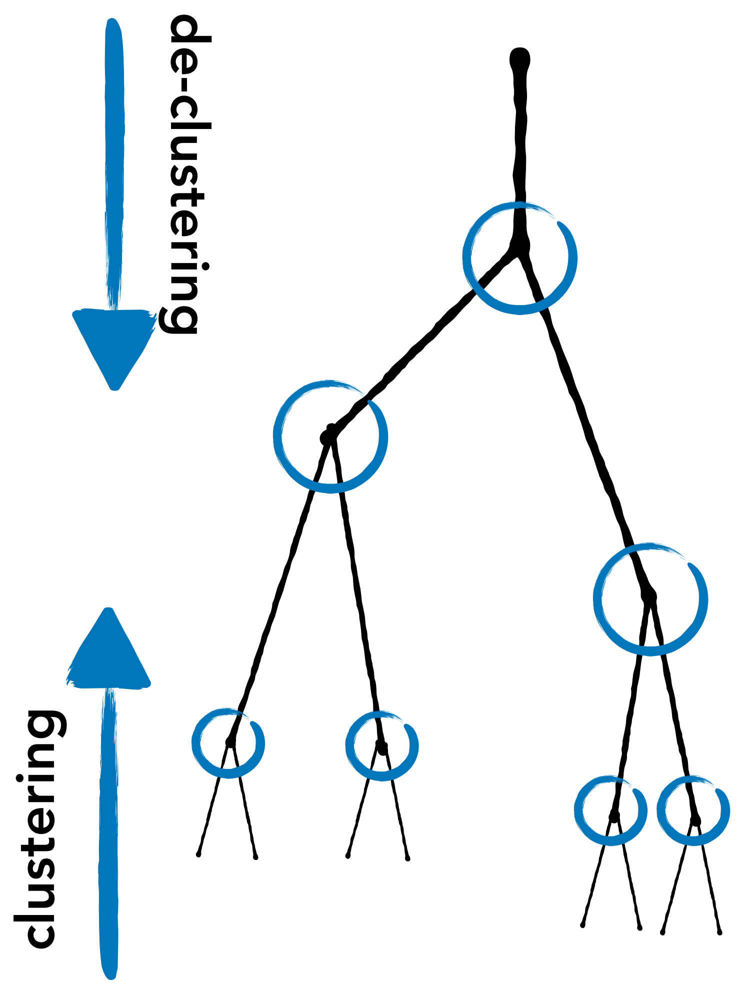
How a jet is formed (& reconstructed)

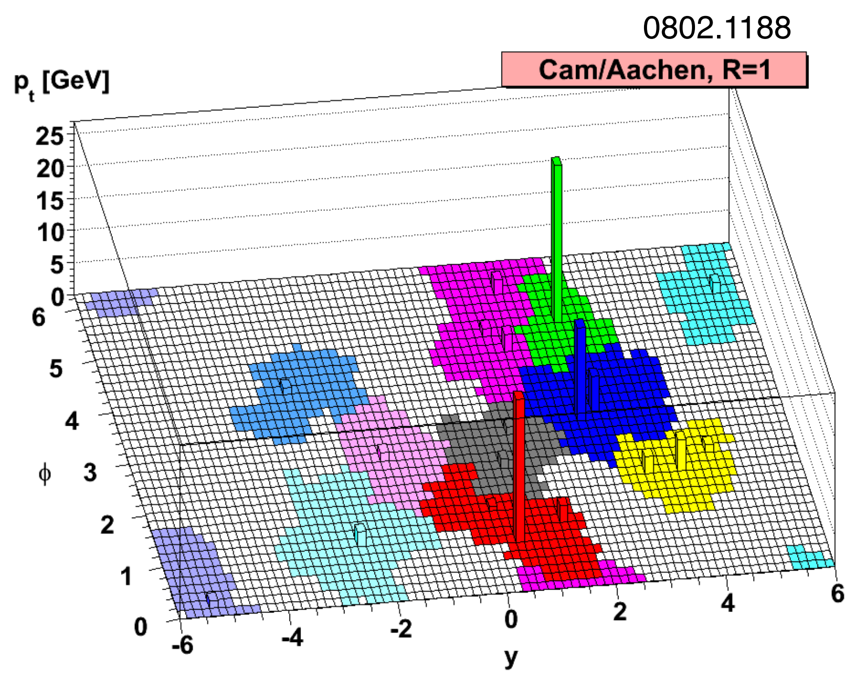
<u>Jet formation</u>: starting from a hard 'seed' - sequential combination of QCD showering, fragmentation and hadronization and possibly massive particle decays



How a jet is formed (& reconstructed)

<u>Jet reconstruction</u>: sequential combination of energy/ momentum (sub) clusters based on appropriate IRC safe measure (e.g. k_T)

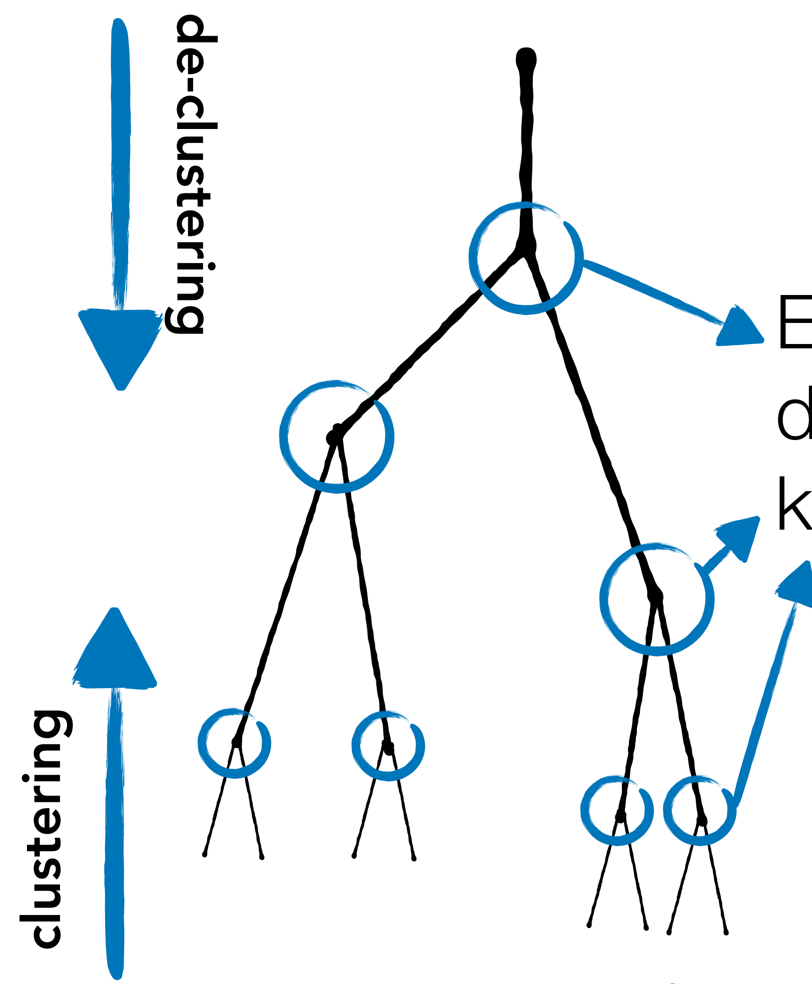




Clustering history - proxy for how jet was formed

How a jet is formed (& reconstructed)

Jet reconstruction: sequential combination of energy/ momentum (sub) clusters based on appropriate IRC safe measure (e.g. k_T)



Each (un)clustering node defines own 'Lund plane' kinematics - observables

$$\Delta \equiv \Delta R_{ij}, \qquad k_t \equiv p_{tj}\Delta, \qquad m^2 \equiv (p_i + p_j)^2$$

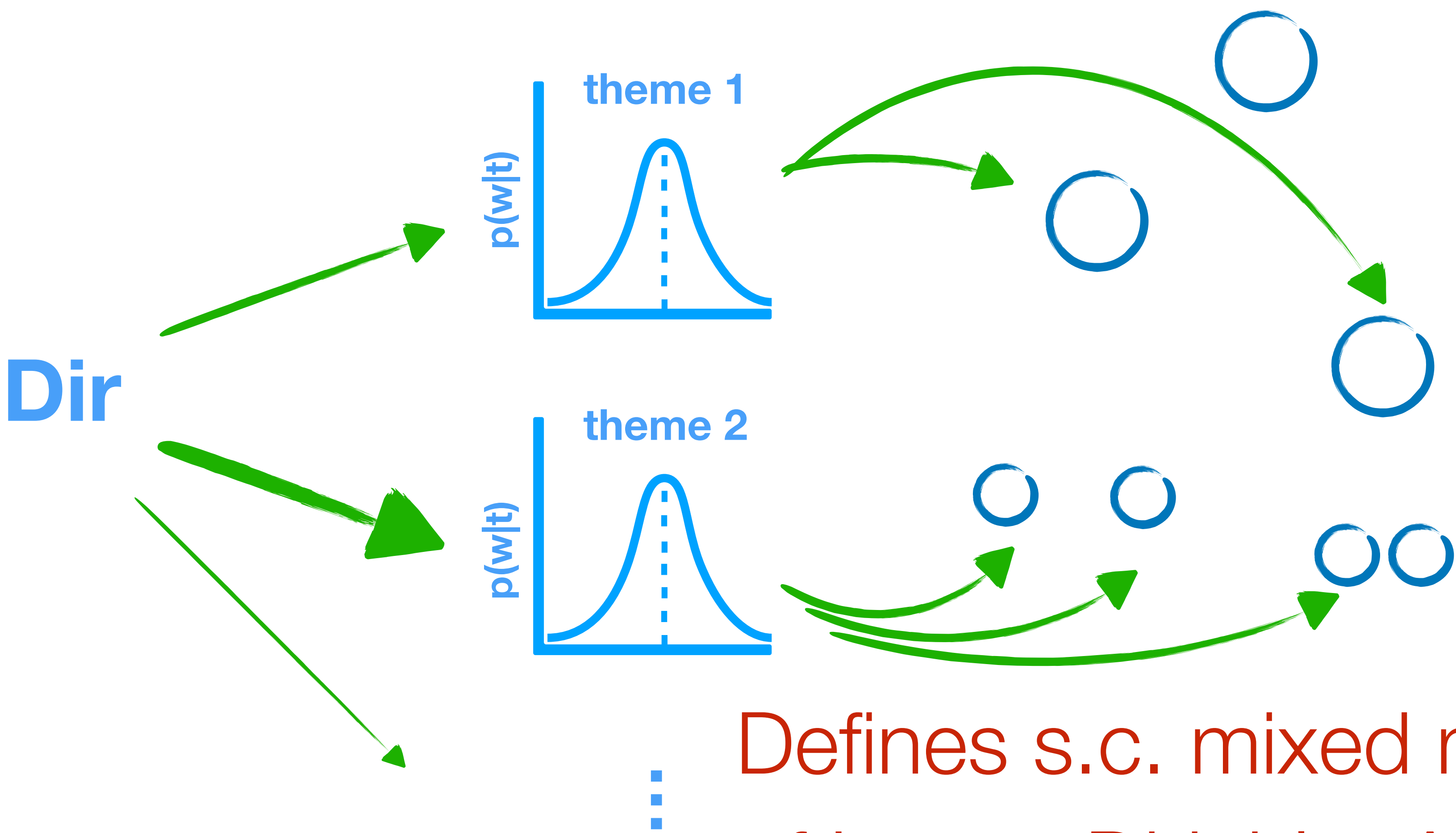
Dreyer, Salam & Soyez, 1807.04758

$$z \equiv \frac{p_{tj}}{p_{ti} + p_{tj}}, \qquad \kappa \equiv z\Delta, \qquad \psi \equiv \tan^{-1}\frac{y_j - y_i}{\phi_j - \phi_i}$$

Simplified generative model of jet (observables)

Assume:

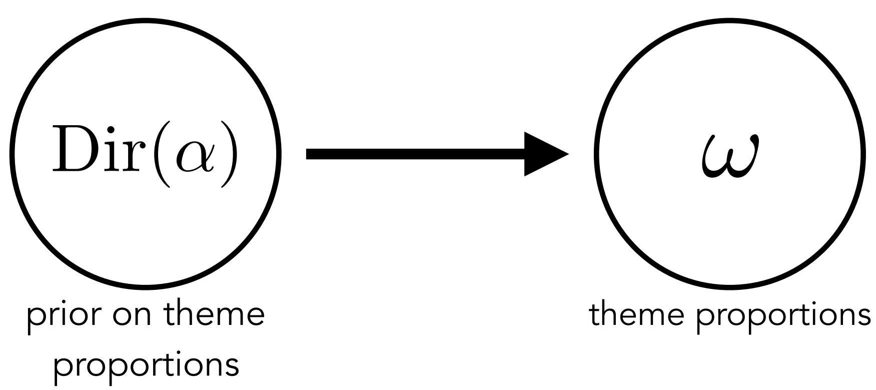
- most of useful jet information contained in node observables
- their values are generated by sampling from several underlying 'latent' distributions (e.g. QCD splitting, particle decay,…) - *themes*



Defines s.c. mixed membership model of Latent Dirichlet Allocation

Latent Dirichlet Allocation for jet observables

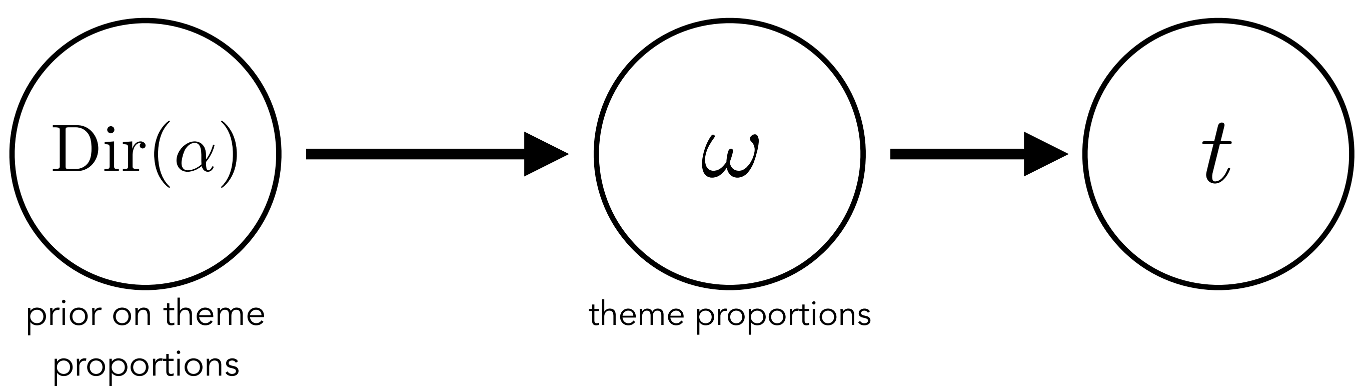
Construct the generative model for jets with K themes



Step 1: sample proportions for each theme, a K-dimensional multinomial

Latent Dirichlet Allocation for jet observables

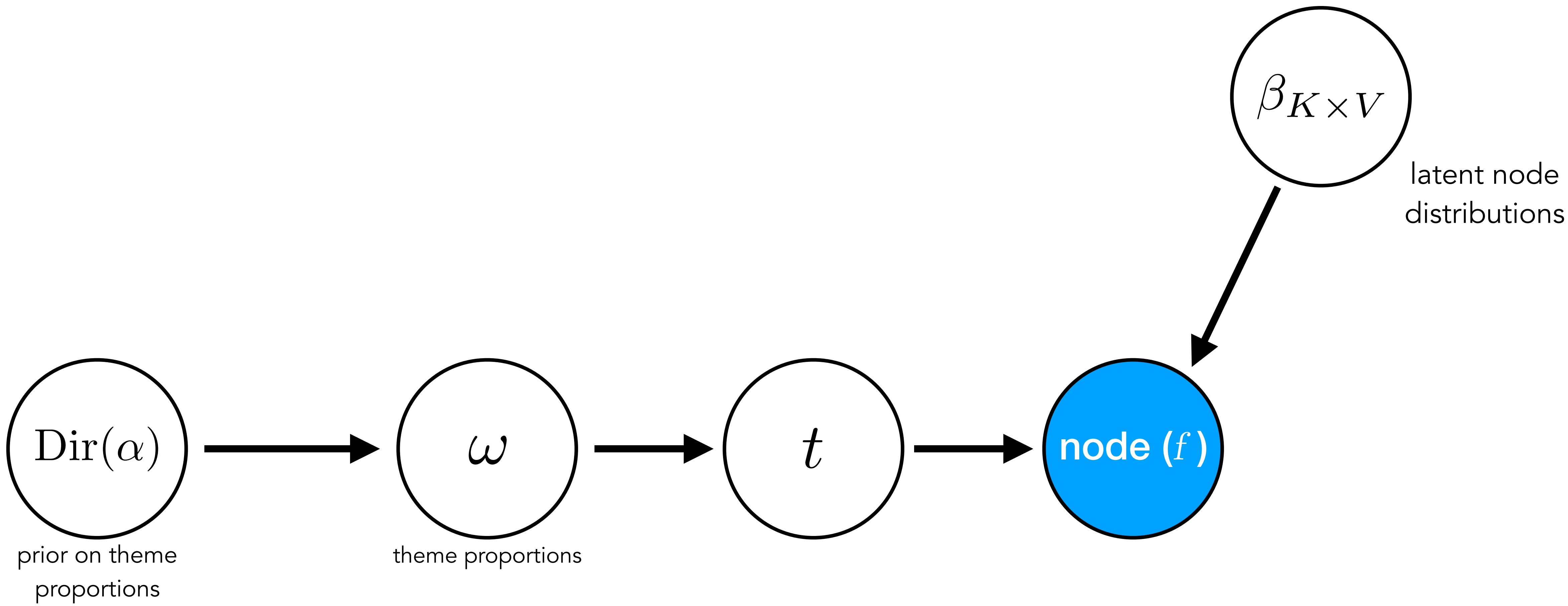
Construct the generative model for jets with K themes



Step 2: sample a single theme from the multinomial

Latent Dirichlet Allocation for jet observables

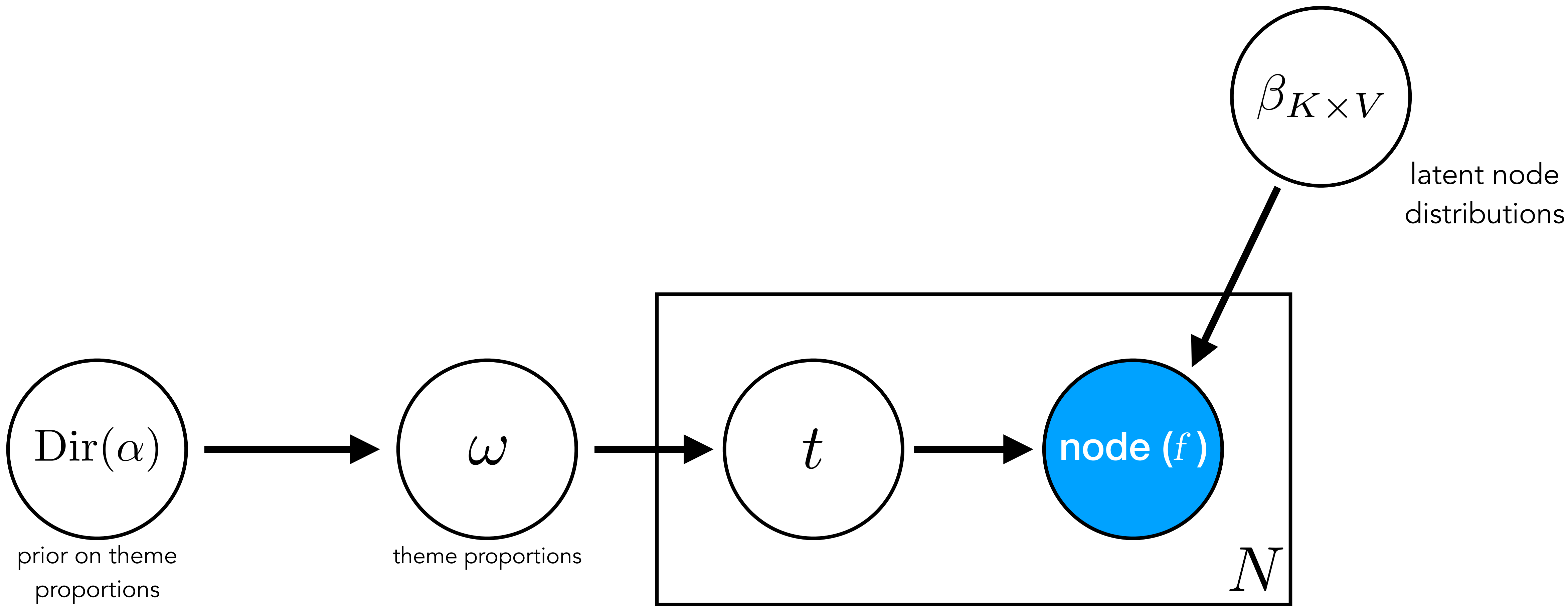
Construct the generative model for jets with K themes



Step 3: sample a node from the appropriate theme distribution

Latent Dirichlet Allocation for jet observables

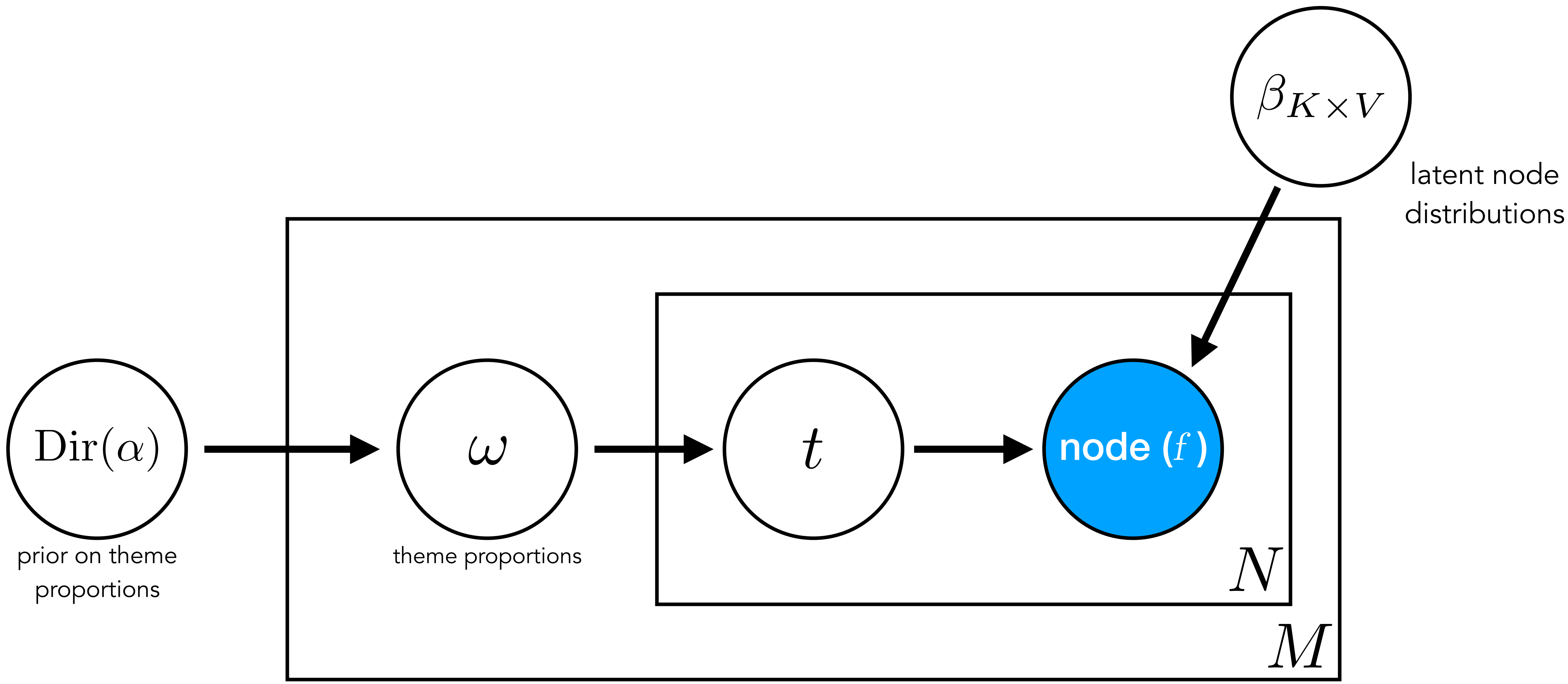
Construct the generative model for jets with K themes



- repeat this for each of the N nodes in the jet

Latent Dirichlet Allocation for jet observables

Construct the generative model for jets with K themes



- repeat this for each of the N nodes in the jet
- repeat again for each of the M jets you want to generate

Latent Dirichlet Allocation for jet observables

Construct the generative model for jets with K themes

Define probability to generate a set of node observables (f)

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

Solve for latent theme distributions (β) using Bayes theorem & approximate inference

$$\beta^{\mathrm{MLE}}_{K\times V} = \underset{\beta}{\mathrm{argmax}} \;\; \log \left(\prod_{i=1}^{M} p\left(\mathrm{jet}_i|\alpha,\beta\right) \right)$$

Originally constructed for study of genotypes & text topics

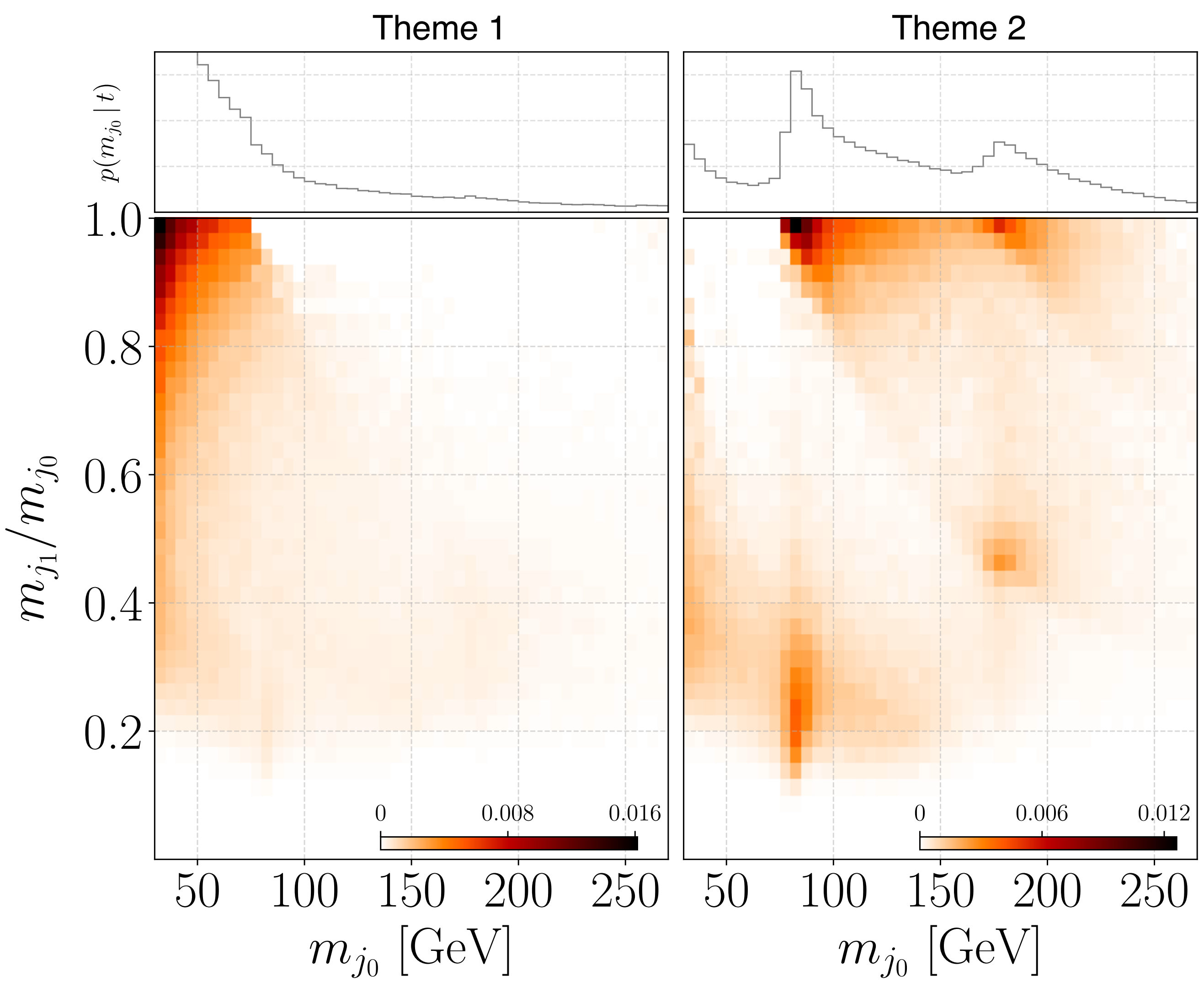
Papadimitriou, Raghavan, Tamaki & Vempala (1998)
Hofmann (1999)
Blei, Ng, & Jordan (2002)

Example LDA models: top jets

2-theme LDA trained on mixed sample (S+B) using jet mass

B: QCD dijets

S: $pp \to t\bar{t} \to W^+W^-b\bar{b}, \quad S/B = 1$

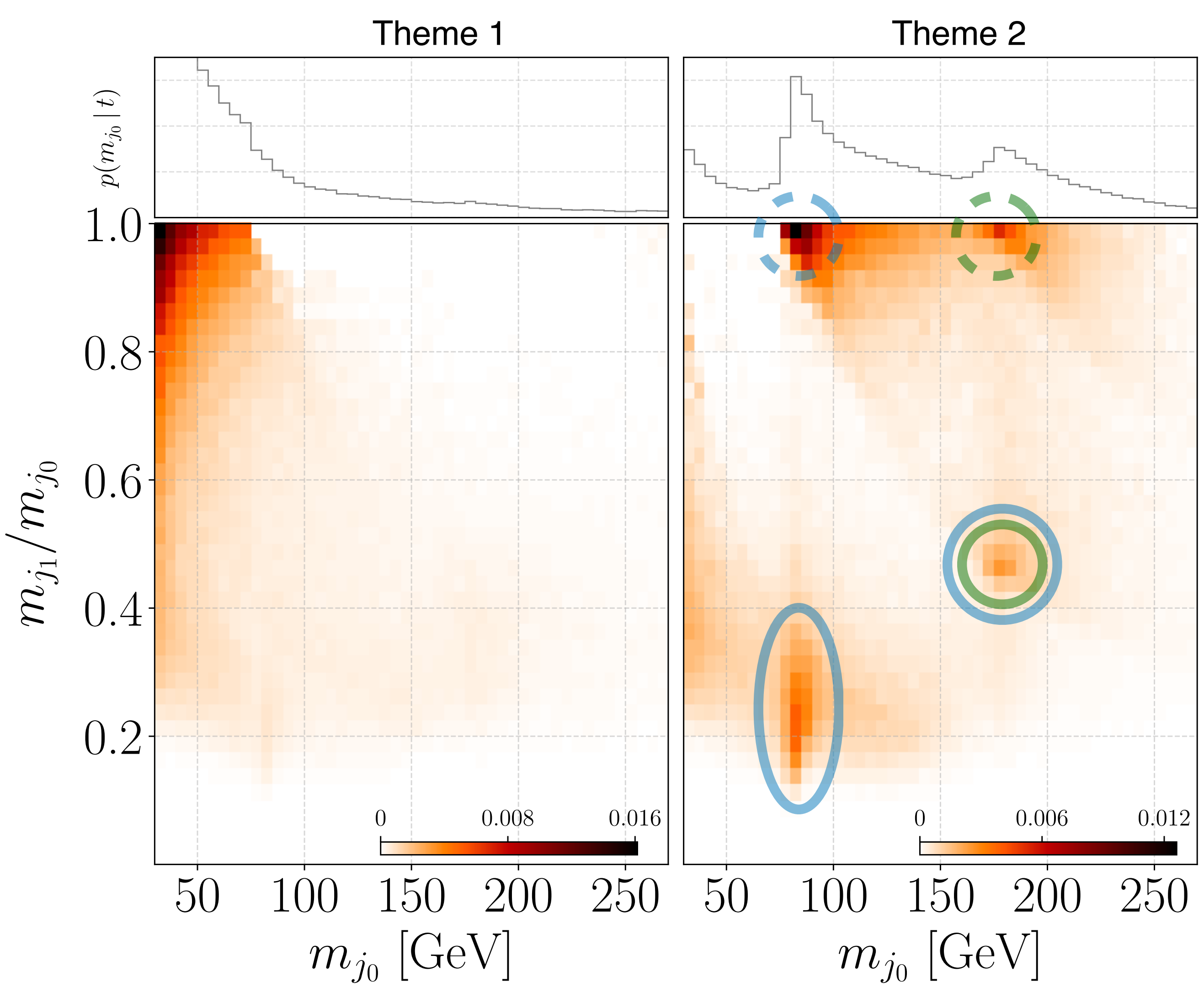


Example LDA models: top jets

2-theme LDA trained on mixed sample (S+B) using jet mass

B: QCD dijets

S: $pp \to t\bar{t} \to W^+W^-b\bar{b}, \quad S/B = 1$



Example LDA models: NP jets

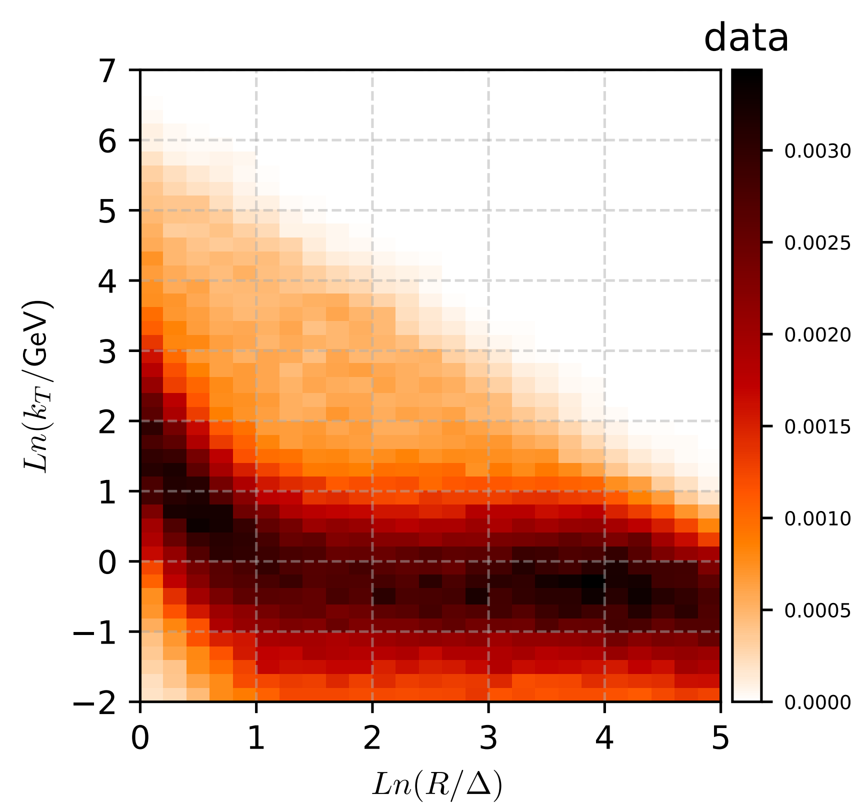
2-theme LDA trained on mixed sample (S+B) in Lund plane

B: QCD dijets

S: $pp \to W' \to W\phi \to WZb\bar{b}$ $\quad S/B = 0.011$

$m_{W'} = 3\ \mathrm{TeV},\ m_\phi = 400\ \mathrm{GeV}$ $\quad m_{jj} \in 2730 - 3190\ \mathrm{GeV}$

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



Example LDA models: NP jets

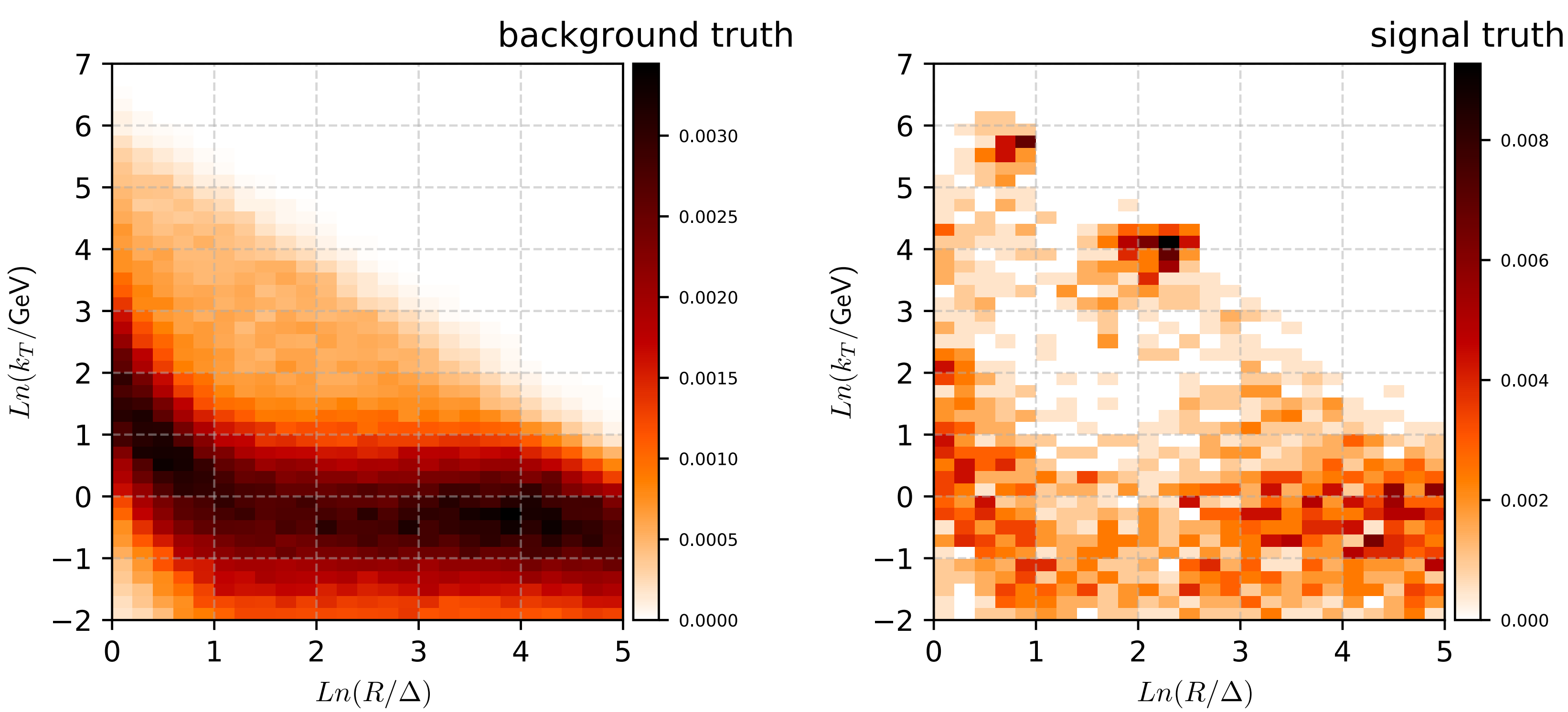
2-theme LDA trained on mixed sample (S+B) in Lund plane

B: QCD dijets

S: $pp \to W' \to W\phi \to WZb\bar{b}$ $\quad S/B = 0.011$

$m_{W'} = 3\ \text{TeV},\ m_\phi = 400\ \text{GeV}$ $\quad m_{jj} \in 2730 - 3190\ \text{GeV}$

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



Example LDA models: NP jets

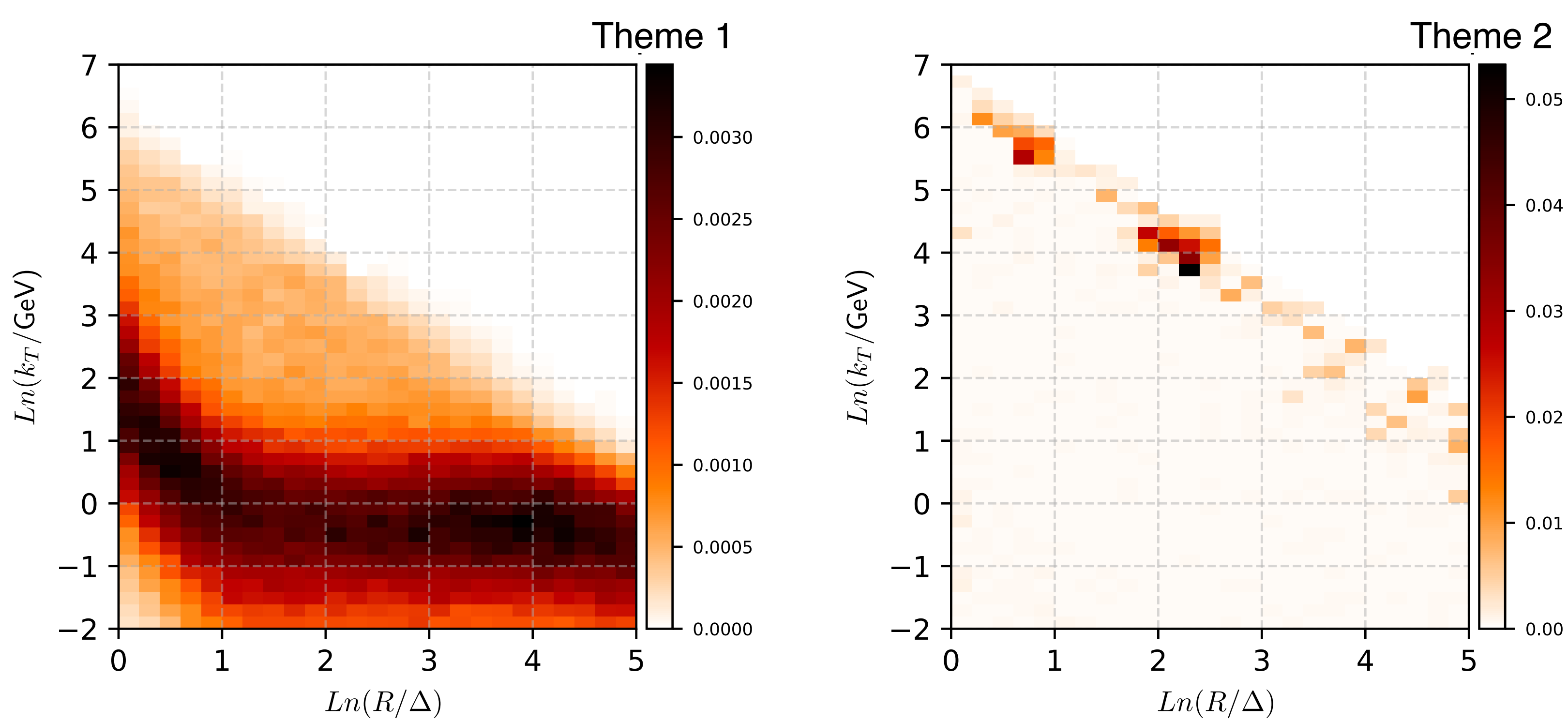
2-theme LDA trained on mixed sample (S+B) in Lund plane

B: QCD dijets

S: $pp \to W' \to W\phi \to WZb\bar{b}$ $\quad S/B = 0.011$

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

$m_{W'} = 3\text{ TeV},\ m_\phi = 400\text{ GeV}$ $\quad m_{jj} \in 2730 - 3190\text{ GeV}$



Example LDA models: NP jets

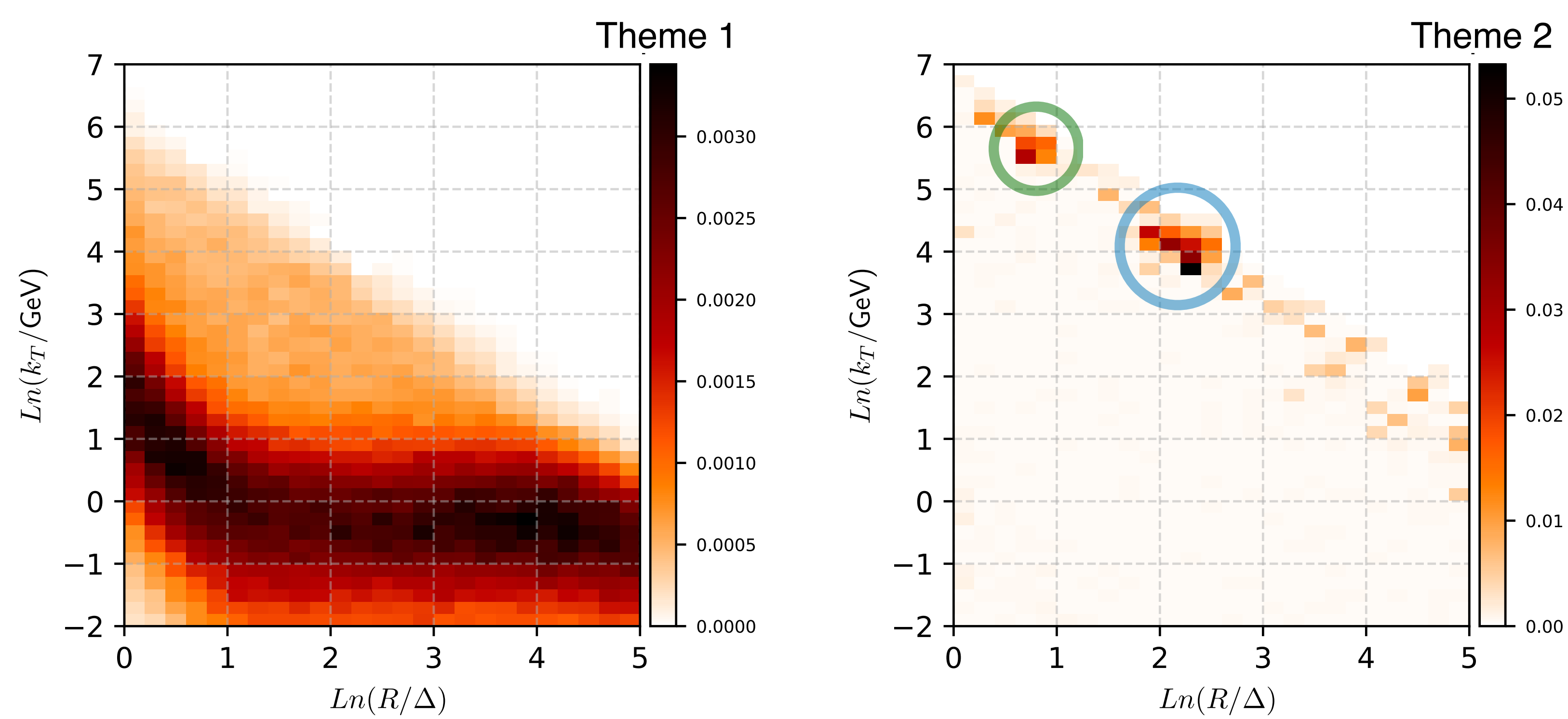
2-theme LDA trained on mixed sample (S+B) in Lund plane

B: QCD dijets

S: $pp \to W' \to W\phi \to WZb\bar{b}$ $\quad S/B = 0.011$

$m_{W'} = 3 \text{ TeV}, \; m_\phi = 400 \text{ GeV} \quad m_{jj} \in 2730 - 3190 \text{ GeV}$

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



Can we use this information to classify QCD from UV?

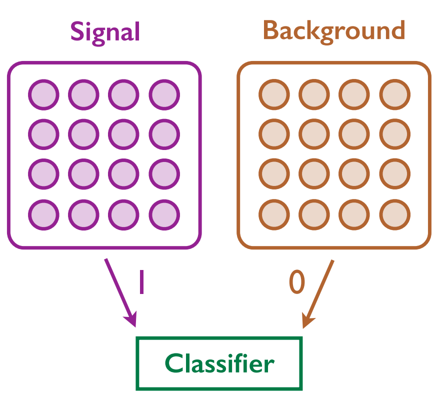
Jet classification: basics

x list of observables useful for distinguishing S from B

$p_S(x)$ and $p_B(x)$ - probability distributions of x for S and B

classifier h(x) close to 1 for S and close to 0 for B - to be learned by minimizing loss function (e.g mean-square)

receiver operating characteristic (ROC) curve



$$\epsilon_S = \int \mathrm{d}\vec{x}\, p_S(\vec{x})\, \Theta(h(\vec{x}) - c)$$
$$\epsilon_B = \int \mathrm{d}\vec{x}\, p_B(\vec{x})\, \Theta(h(\vec{x}) - c)$$

Neyman-Pearson lemma: $h_{\mathrm{optimal}}(\vec{x}) = p_S(\vec{x})/p_B(\vec{x})$ (likelihood ratio)

If x - low dimensional, can use histograms directly, otherwise use supervised ML (BDTs, NNs, …)

Jet classification: basics

x list of c

$p_S(x)$ anc

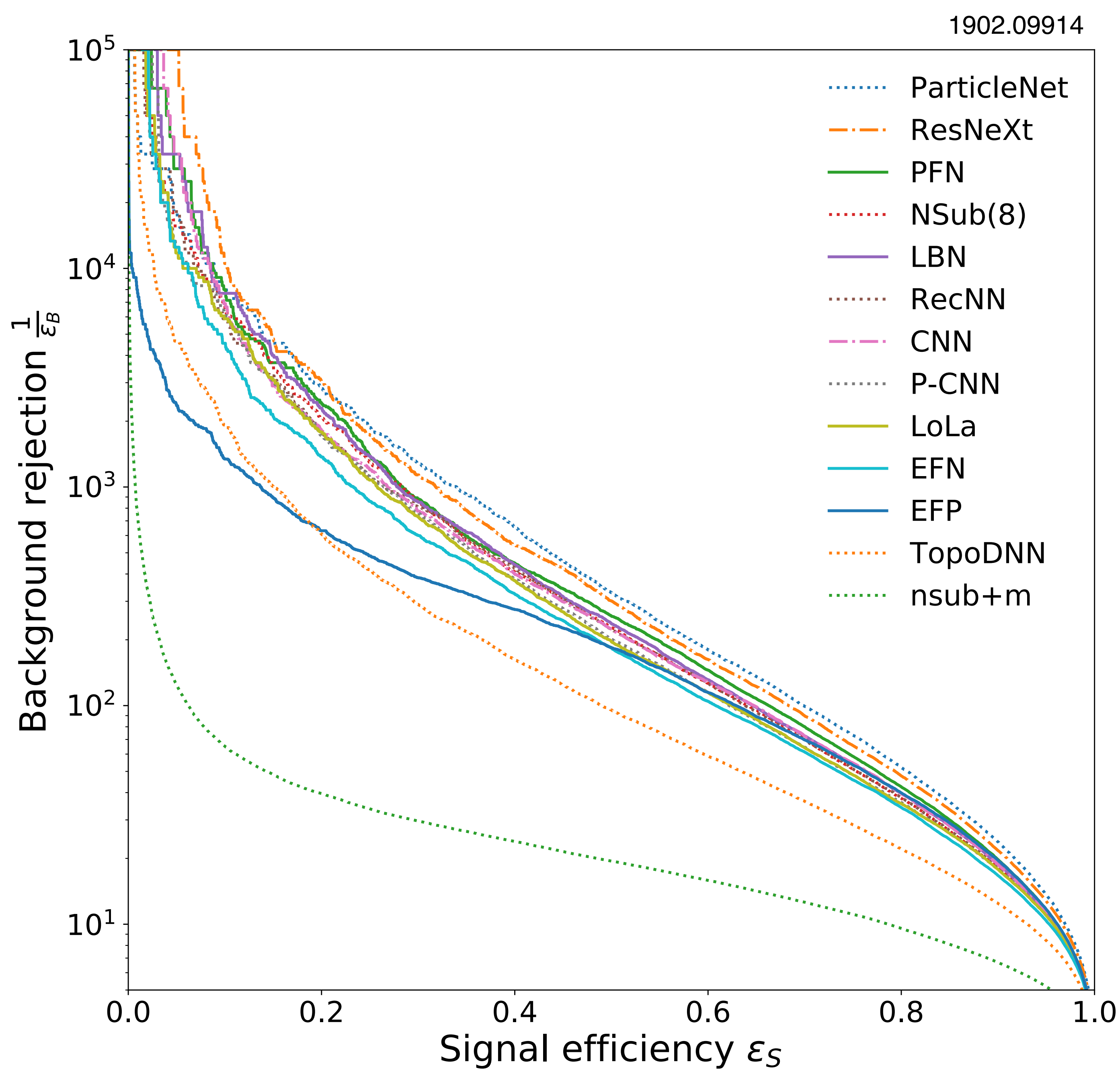
classifier
learned b

receiver

Neyman-

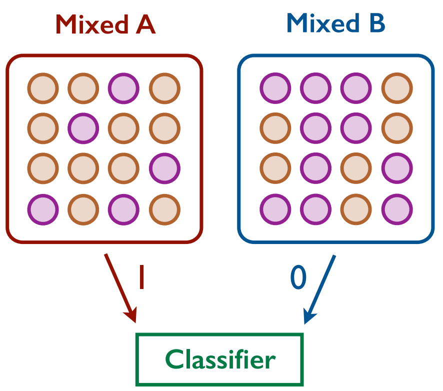
If x - low
otherwis

Example: QCD j vs. t classification (top-tagging)



Jet classification: mixed samples

Classification from mixed samples: pure samples not available in real data



$$p_{M_1}(\vec{x}) = f_1\, p_S(\vec{x}) + (1 - f_1)\, p_B(\vec{x}),$$
$$p_{M_2}(\vec{x}) = f_2\, p_S(\vec{x}) + (1 - f_2)\, p_B(\vec{x}),$$

1.) Assume f_1, f_2 known (e.g. from MC), then simply

$$h^{M_1/M_2}_{\text{optimal}}(\vec{x}) = p_{M_1}(\vec{x}) / p_{M_2}(\vec{x})$$

2.) Assume only $f_1 > f_2$ then use monotonicity of

$$\frac{p_{M_1}}{p_{M_2}} = \frac{f_1\, p_S + (1 - f_1)\, p_B}{f_2\, p_S + (1 - f_2)\, p_B}$$

(Classification Without Labels)

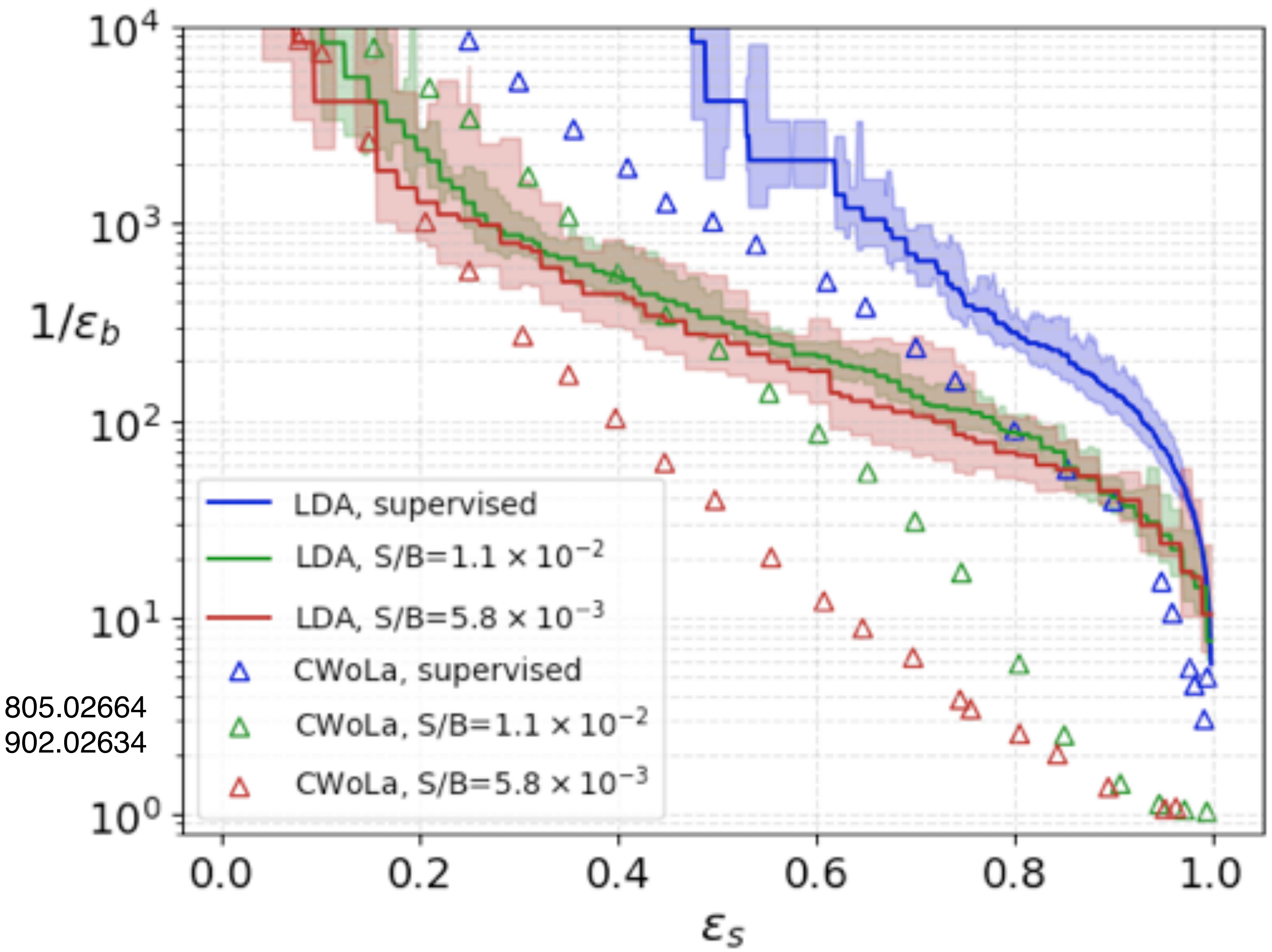
Metodiev, Nachman & Thaler, 1708.02949

Can be used directly on latent theme distributions!
(alternatively use inference on ω)

Jet classification with LDA

NP example: $pp \to W' \to W\phi \to WZb\bar{b}$

$m_{W'} = 3\ \mathrm{TeV},\ m_\phi = 400\ \mathrm{GeV}$ $\qquad m_{jj} \in 2730 - 3190\ \mathrm{GeV}$



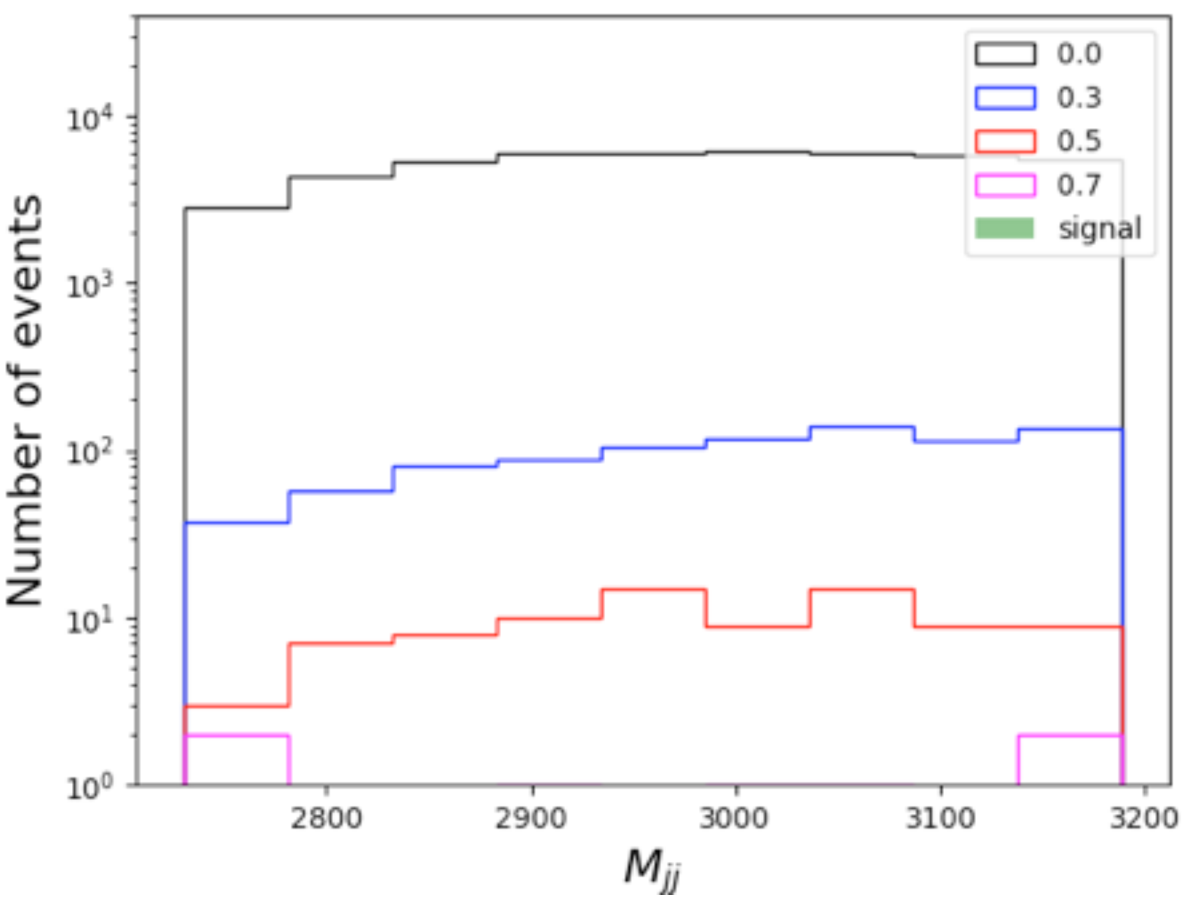
Collins, Howe & Nachman, 1805.02664
1902.02634

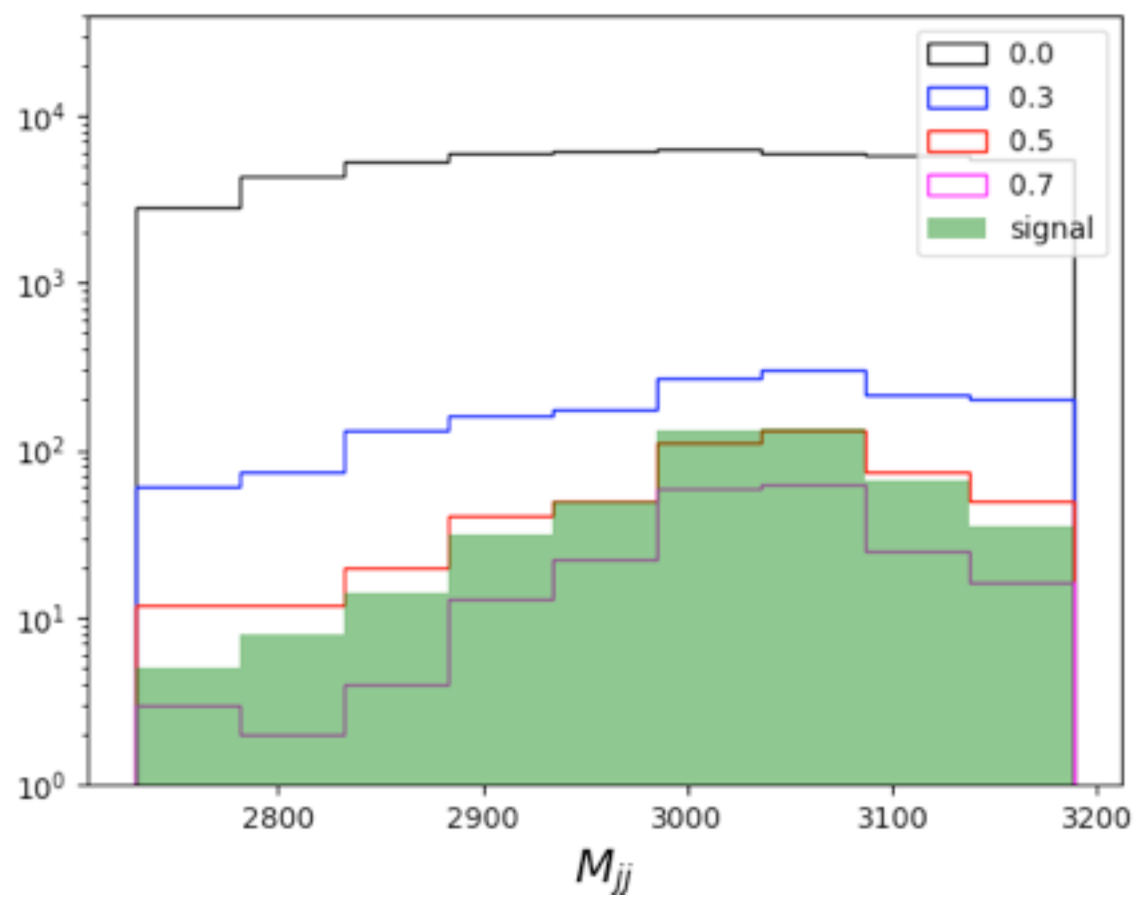
errors estimated using k-folding, with k=10

Bump hunting with LDA

NP example: $pp \to W' \to W\phi \to WZb\bar{b}$

$m_{W'} = 3\ \mathrm{TeV},\ m_\phi = 400\ \mathrm{GeV} \qquad m_{jj} \in 2730 - 3190\ \mathrm{GeV}$





LDA with no signal vs LDA with signal

Conclusions

Latent Dirichlet Allocation addresses several interesting problems in NP (QCD) vs NP (UV) classification:

- ✔ train on mixed, unlabelled, imbalanced samples
- ✔ sensitivity to small S/B (~1%)
- ✔ extract descriptions of signal and background
- ✔ interpret 'what the machine has learned'
- ✔ no control/side-band regions

So far, this is proof-of-concept.

Next steps:

- ❋ explore classifiers with more than two themes
- ❋ hierarchical theme models: extract optimal #themes from data
- ❋ can observables' space be optimised with deep learning?
- ❋ more applications: category based searches, pile-up mitigation, …

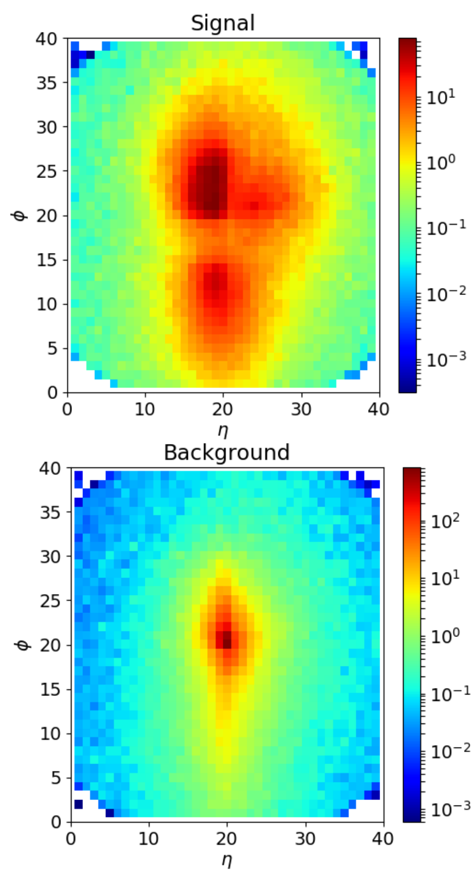
Spares

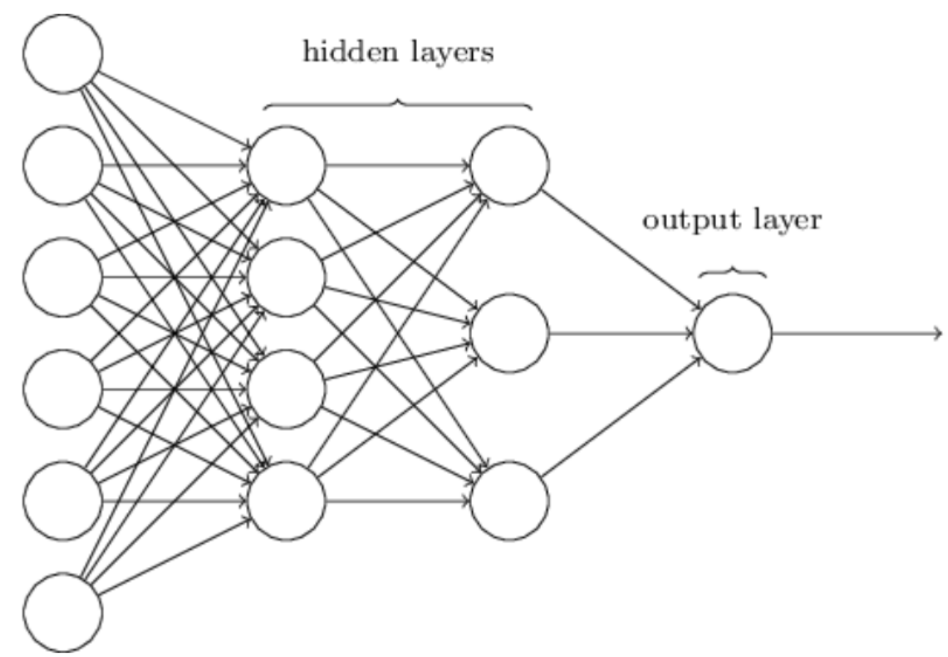
Jet tagging

Recent ML revolution:

see 1902.09914 for recent review

– train NNs on (low level) jet info (jet images, particle four vectors) using (MC generated) pure S & B samples



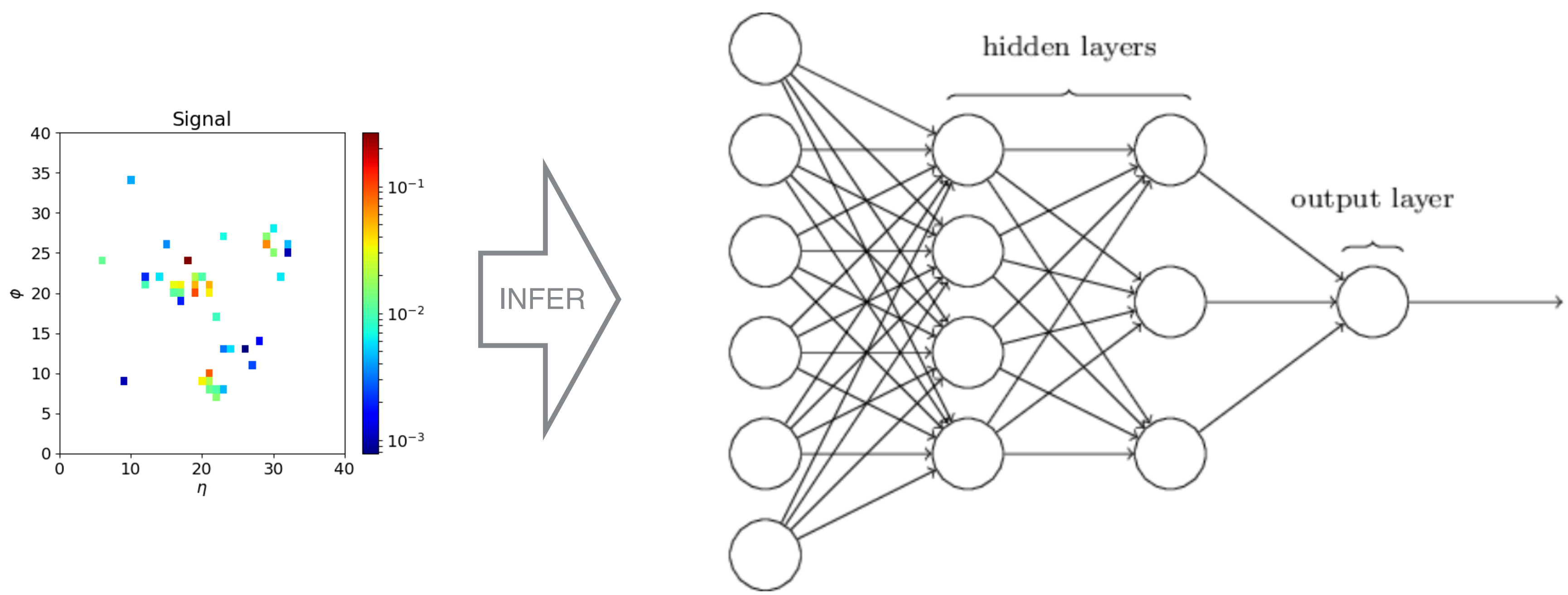


Jet tagging

Recent ML revolution:

see 1902.09914 for recent review

– train NNs on (low level) jet info (jet images, particle four vectors) using (MC generated) pure S & B samples

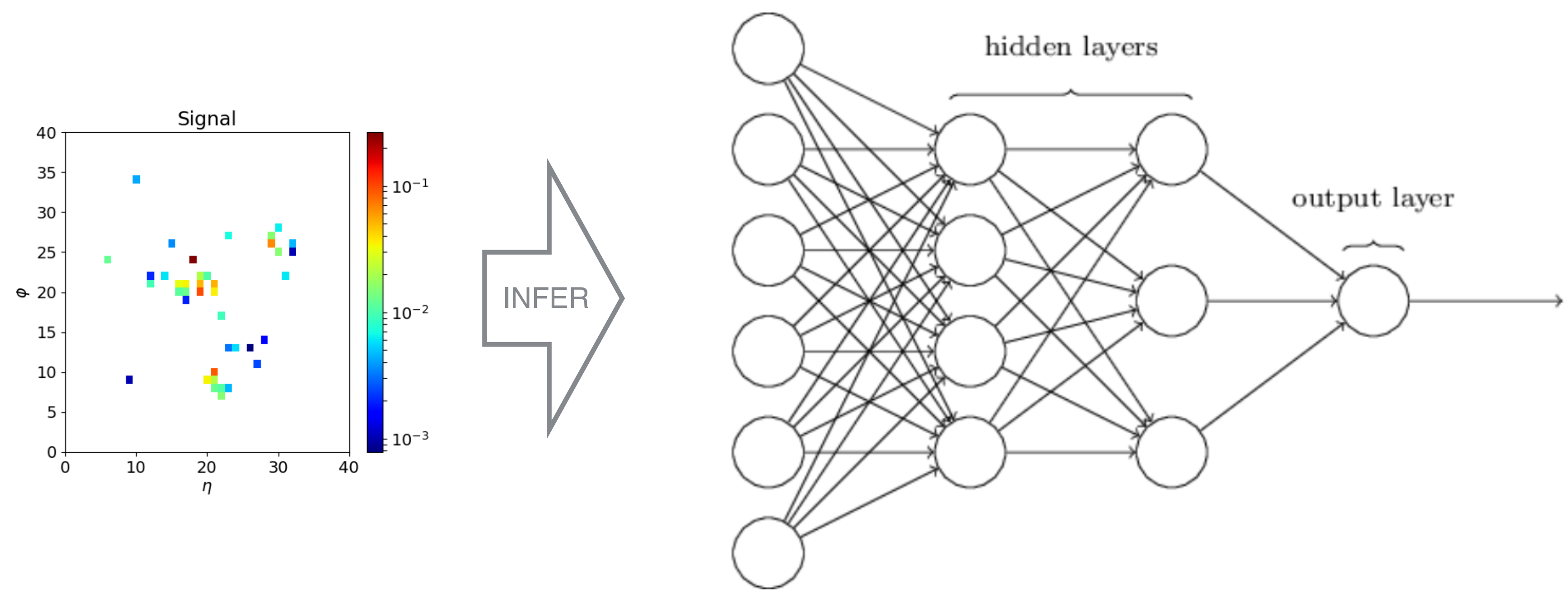


Jet tagging

Recent ML revolution:

see 1902.09914 for recent review

– train NNs on (low level) jet info (jet images, particle four vectors) using (MC generated) pure S & B samples



Key challenge: knowing what the machine has learned (genuine SD physics or MC implementation particularities)

What exactly is the Dirichlet Distribution

Multivariate equivalent of Beta distribution
(e.g. dice factory vs. coin factory)

$$p(\theta|\alpha) = \frac{\Gamma(\sum_{i=1}^{k} \alpha_i)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \prod_{i=1}^{k} \theta_i^{\alpha_i - 1}$$

α_i determines prior - mean shape and sparsity

Dirichlet is defined over (k-1) simplex (k non-negative arguments which sum to one)

Dirichlet is conjugate prior to multinomial distribution - posterior is also Dirichlet

In jet LDA, themes are V-dimentional Dirichlet; theme proportions are K-dimensional Dirichlet

What exactly is the Dirichlet Distribution

Multivariate equivalent of Beta distribution
(e.g. dice factory vs. coin factory)

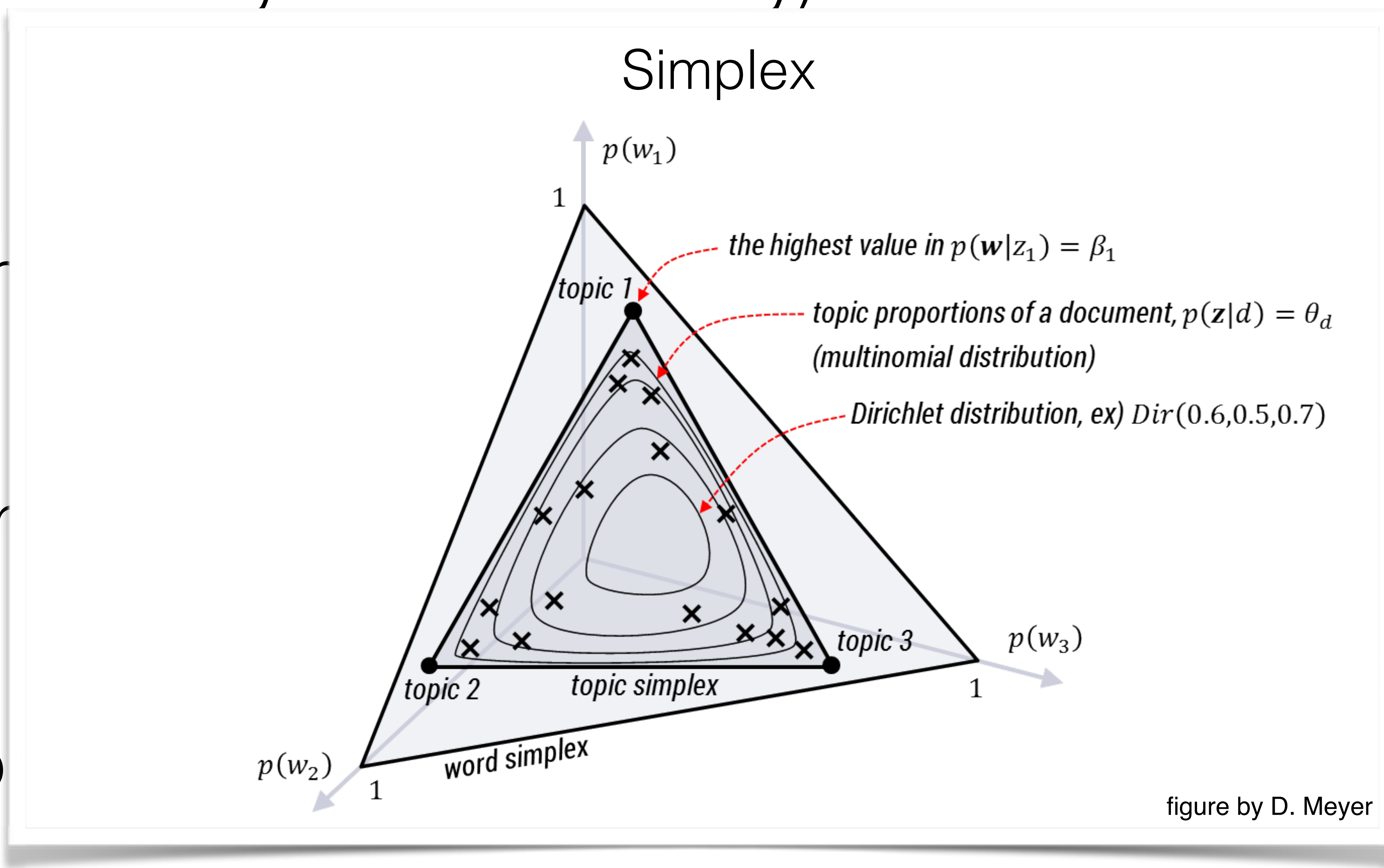


figure by D. Meyer

α_i deterr

Dirichlet /e
argumer

Dirichlet n -
posterio

In jet LDA, themes are V-dimentional Dirichlet; theme proportions are K-dimensional Dirichlet