Shower Deconstruction

Davison E. Soper
University of Oregon

Work with Michael Spannowsky

Florence, September 2011

Introduction

- One can examine the substructure of jets to dig out new physics signals.
- I take a signal event to be one in which one or more new heavy particles is created and decays.
- Sometimes there is a chain of decays.
- It is often useful to arrange that the heavy particle is highly boosted in the transverse direction, even though this may cost in cross section.
- Methods include "mass drop + filtering," "trimming," and "pruning."

- Michael Spannowsky and I propose a general method for subjet analysis: “shower deconstruction.”
- This work originated at the Northwest Terascale workshop on jet substructure held in 2009 at the University of Washington, organized by Steve Ellis.
- A lot of the structure of this comes from the partitioned dipole shower algorithms (now being turned into a parton shower event generator) by Zoltan Nagy and me.
- Our example is finding a Higgs boson that recoils against a Z-boson.

Our example of a signal

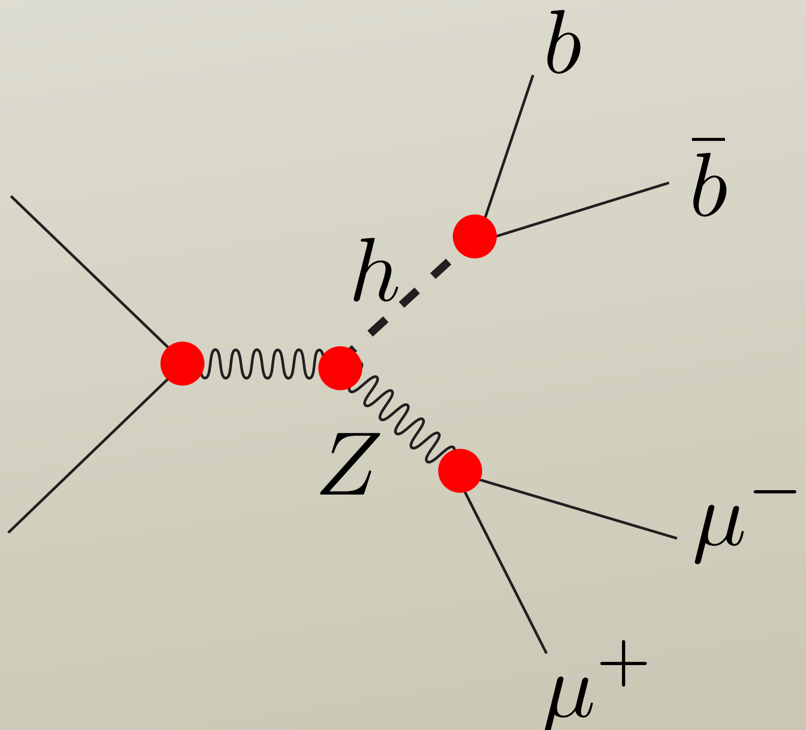
- $p + p \to h + Z$

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

$$h \to b + \bar{b}$$

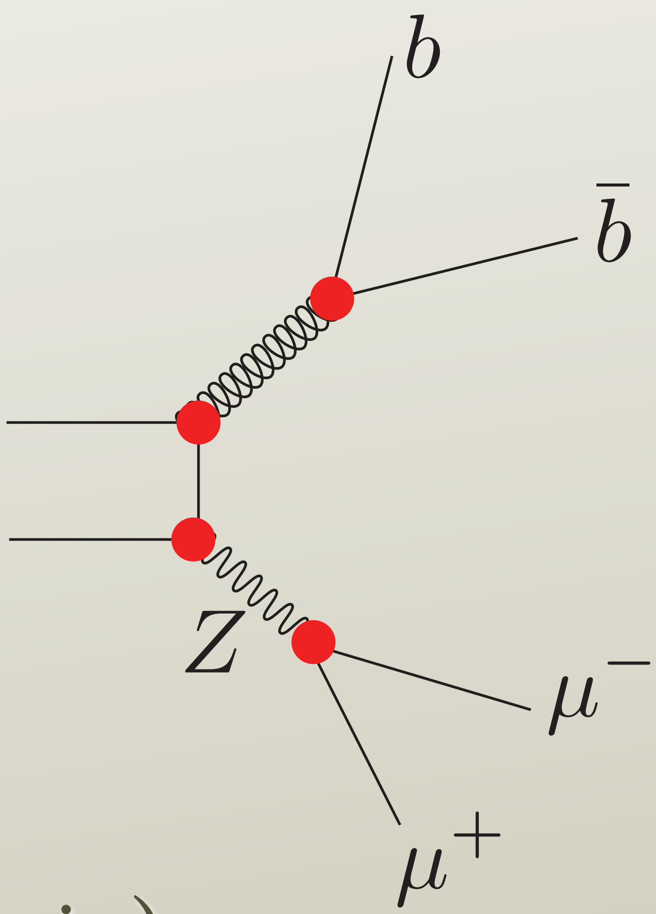
- The Z has lots of p_T so the h is highly boosted.

A Higgs boson signal event



Background events

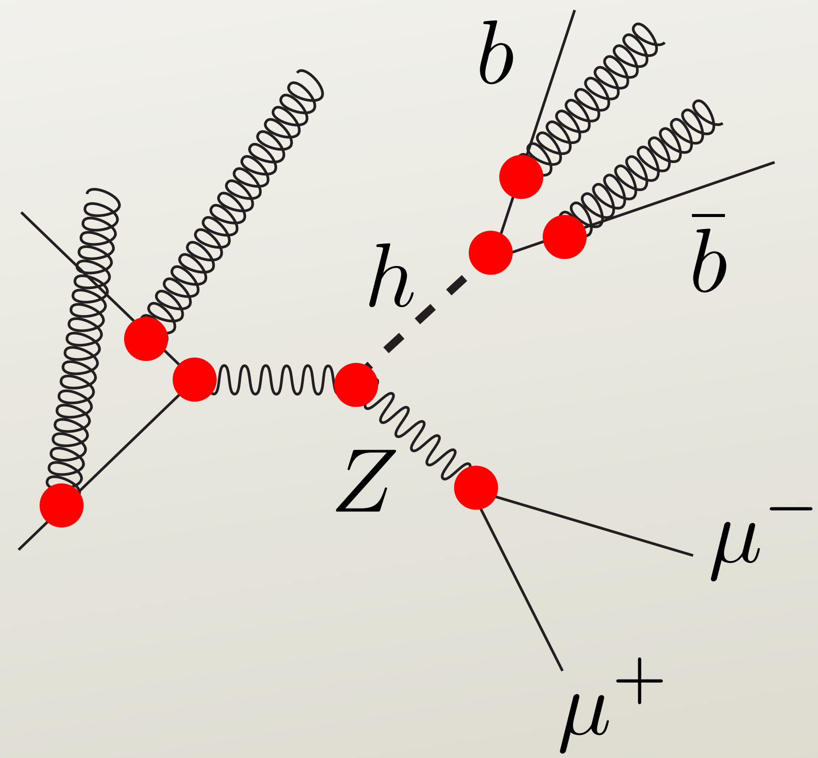
- We need to be able to tell the signal events from QCD background events.



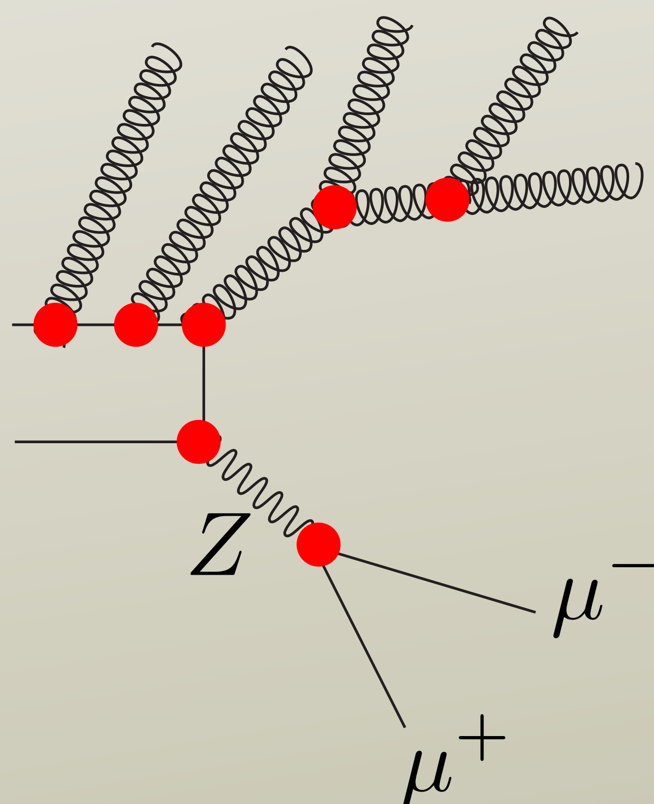
...(at least on a statistical basis.)

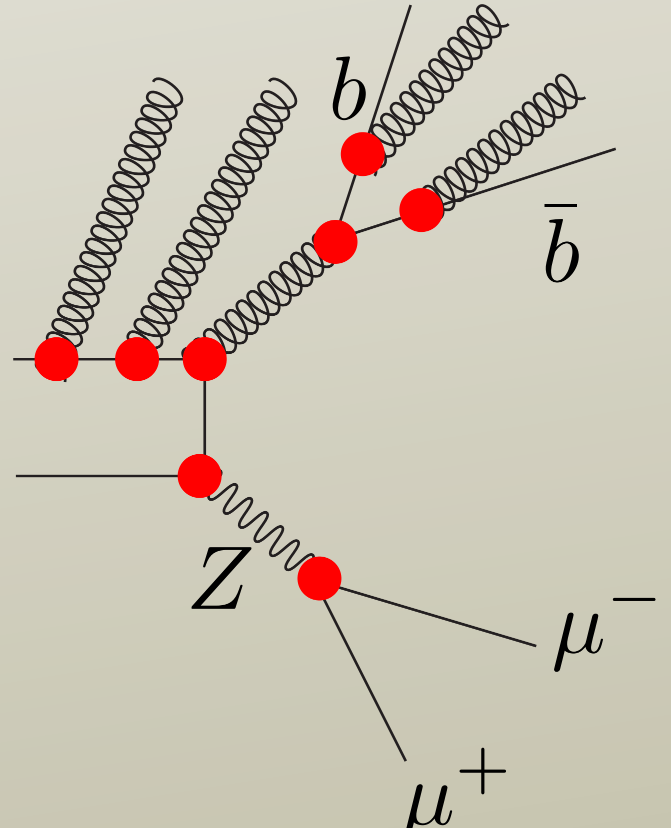
Reality is a bit worse...

We want to find this.



In a background of this.

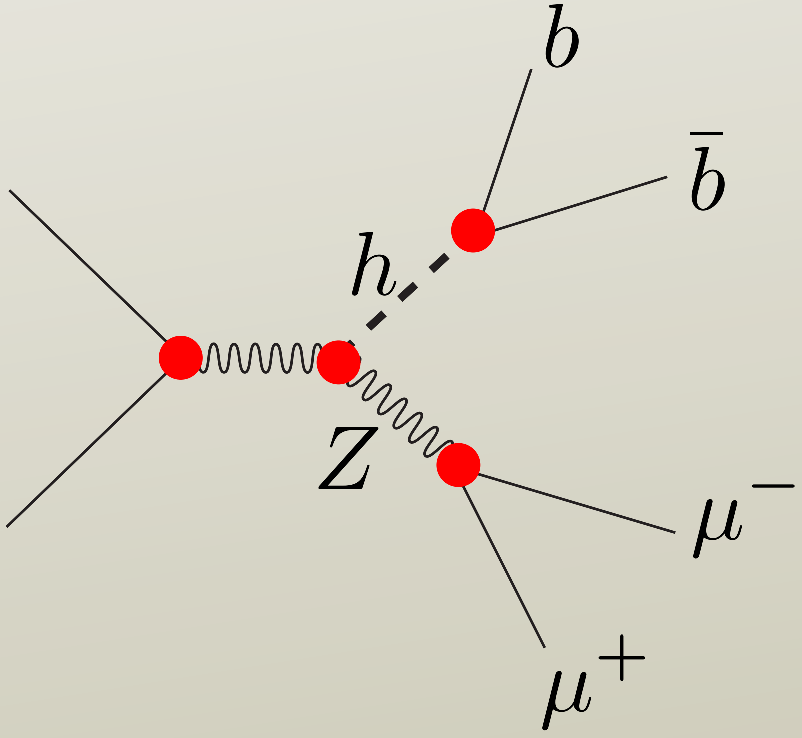




The method of BDRS

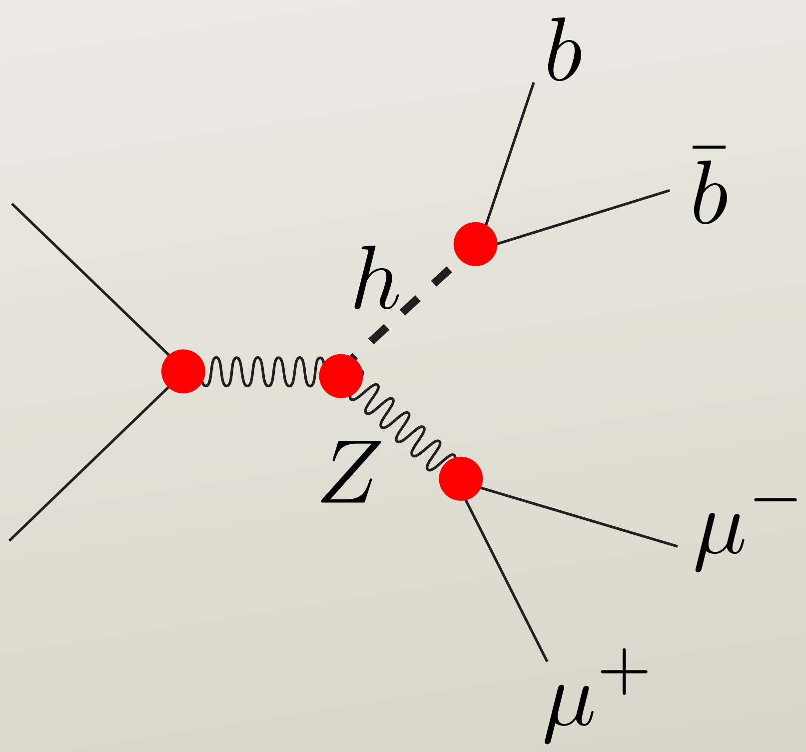
Butterworth, Davison, Rubin, and Salam (2008)

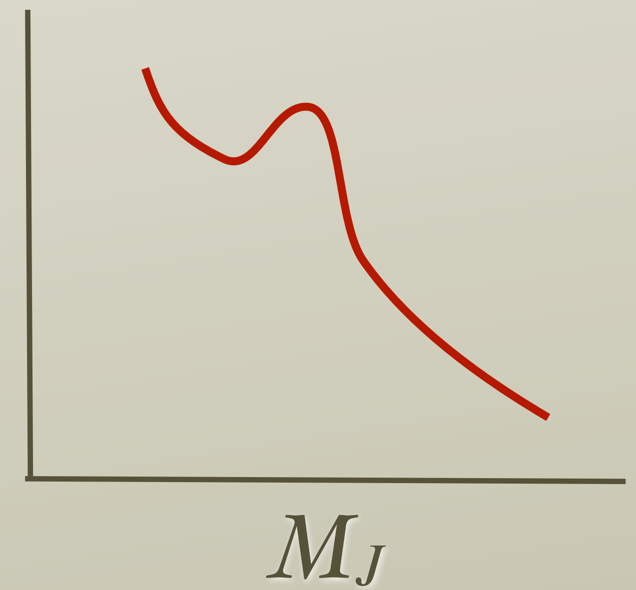
- Electron or muon pair near the Z mass.
- Large P_T of the lepton pair and of recoiling jet (>200 GeV).
- Large P_T implies that the possible Higgs decay products are easier to isolate: they are part of a (rather fat) jet.

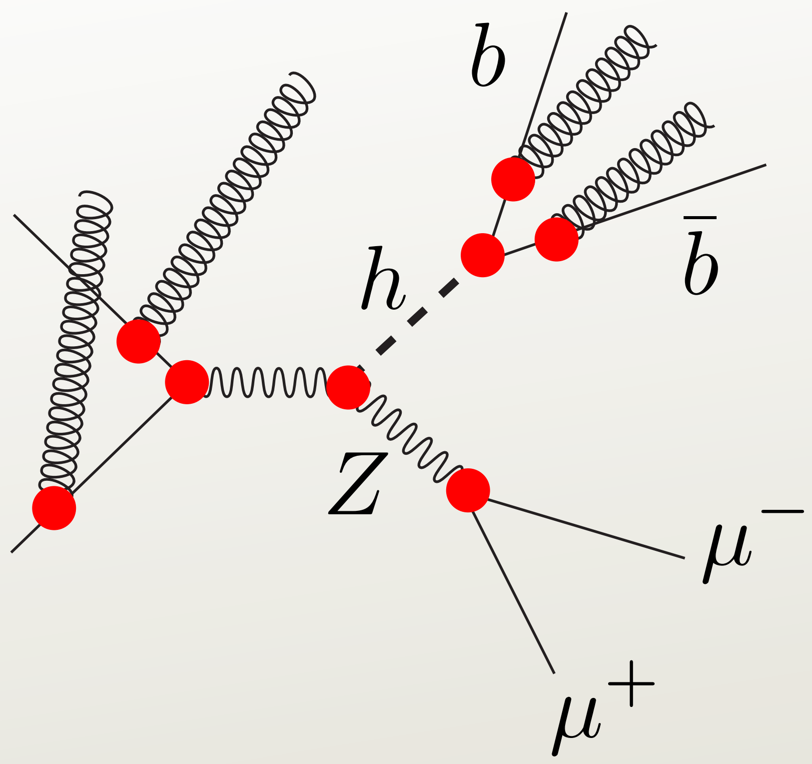


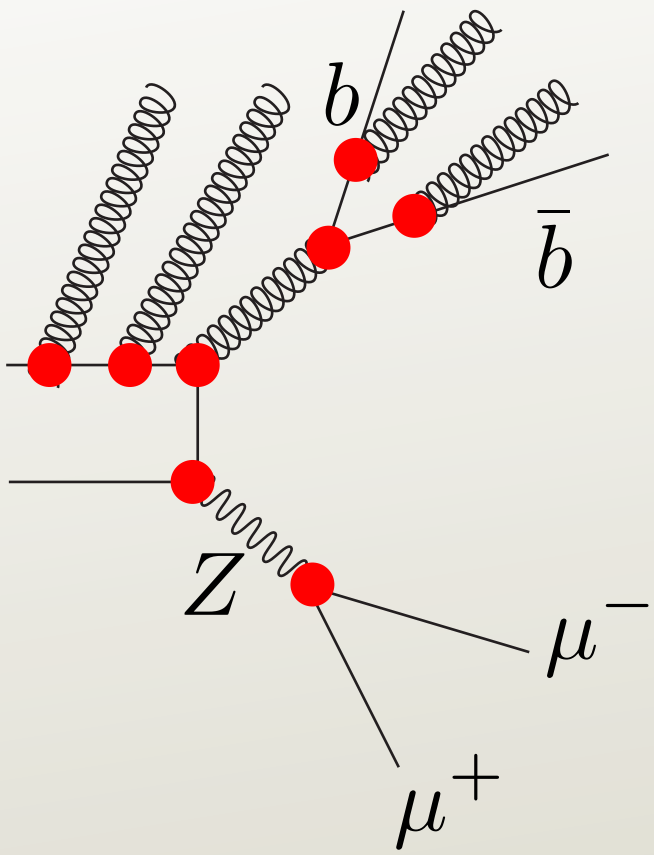
Define the fat jet

- Look for a high P_T jet using the Cambridge-Aachen (angle) algorithm with R=1.2.
- We might hope that the distribution of the mass of the fat jet shows a bump at the Higgs mass.

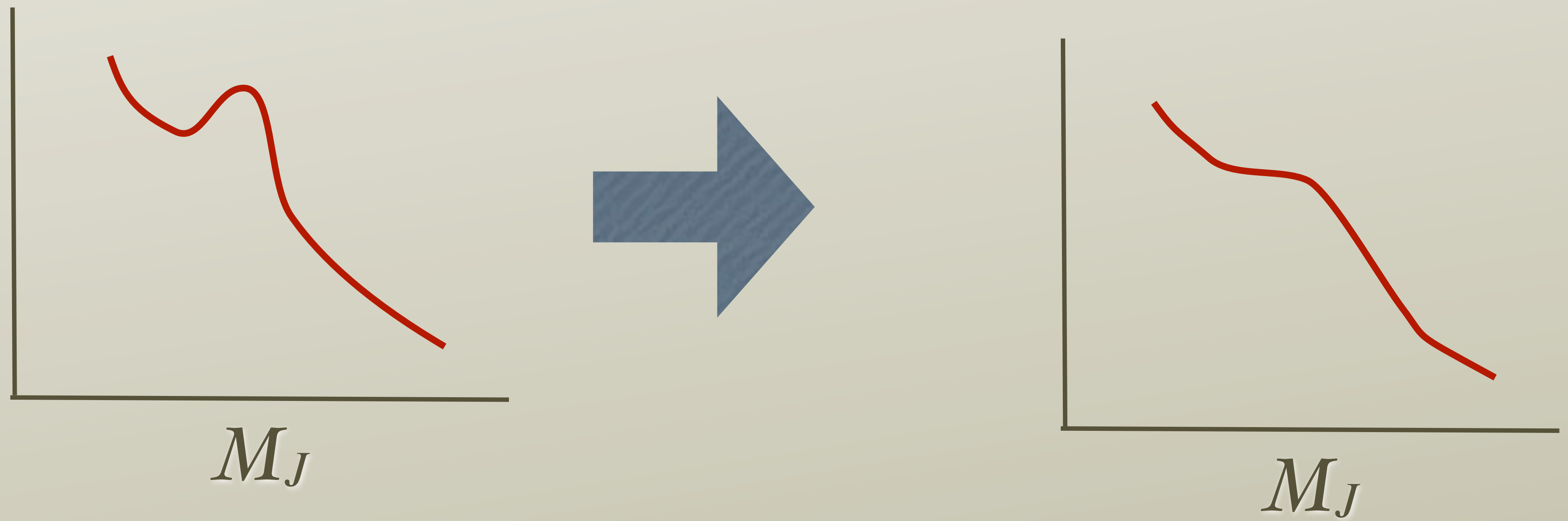






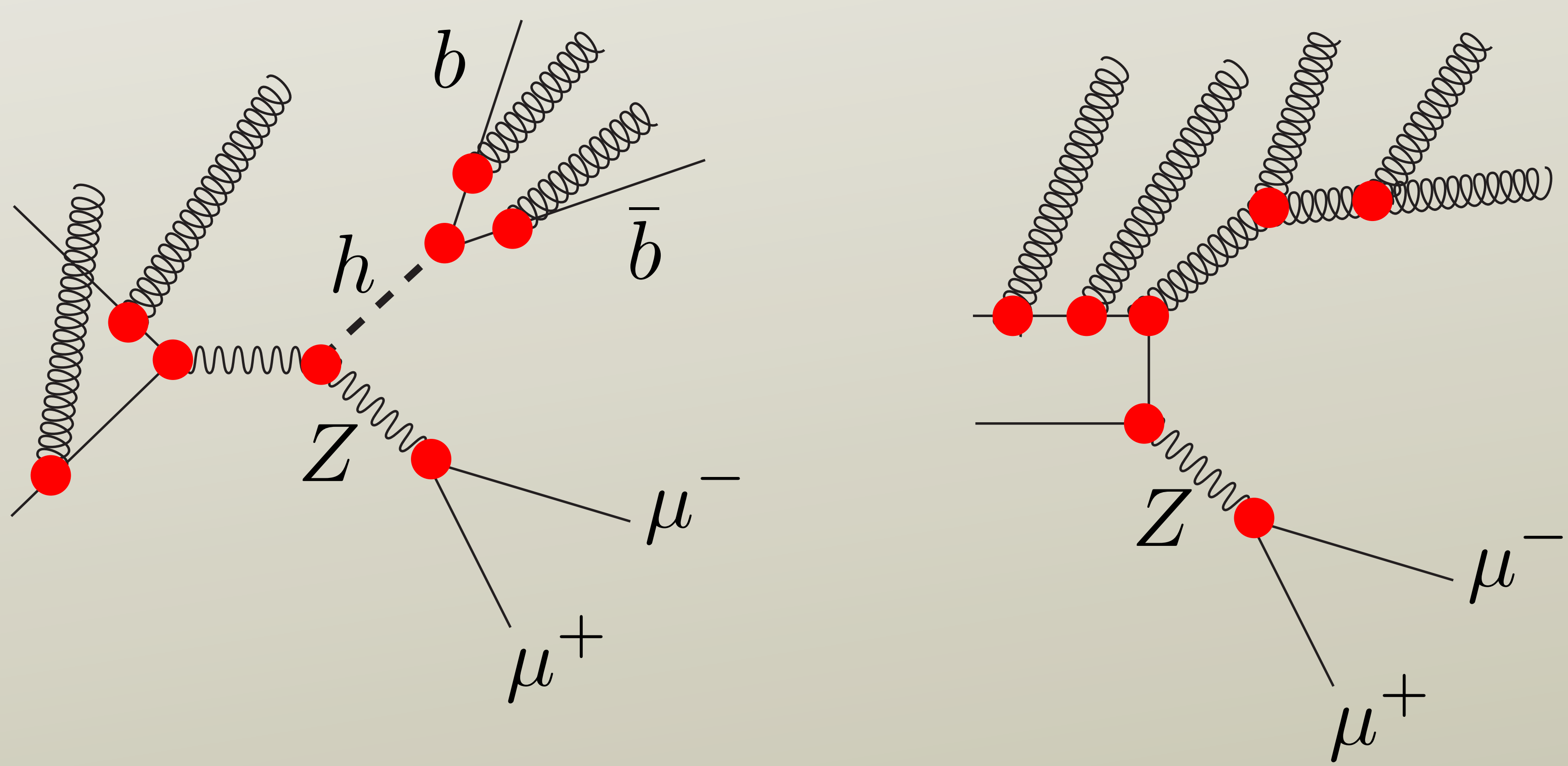


- Since QCD is operating, the mass bump gets smeared out.



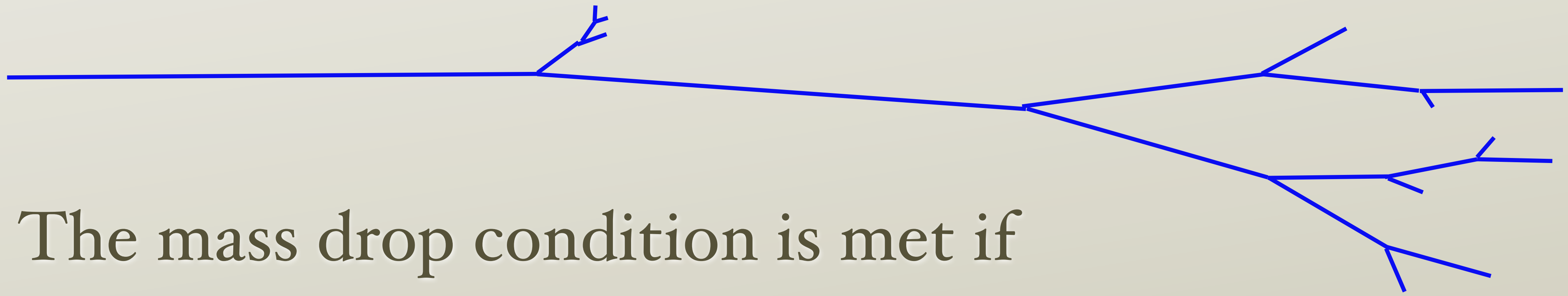
Subjet analysis

- We would like to take apart the fat jet in order to get rid of the contaminating initial state radiation.



Jet mass drop and filtering

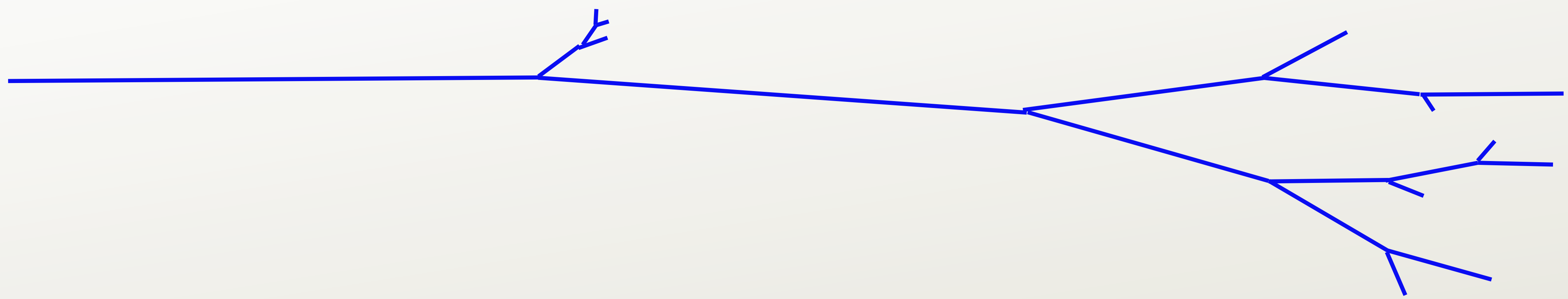
- Step I: mass drop.
  - Examine the C-A splitting tree, starting at the trunk.



  - The mass drop condition is met if

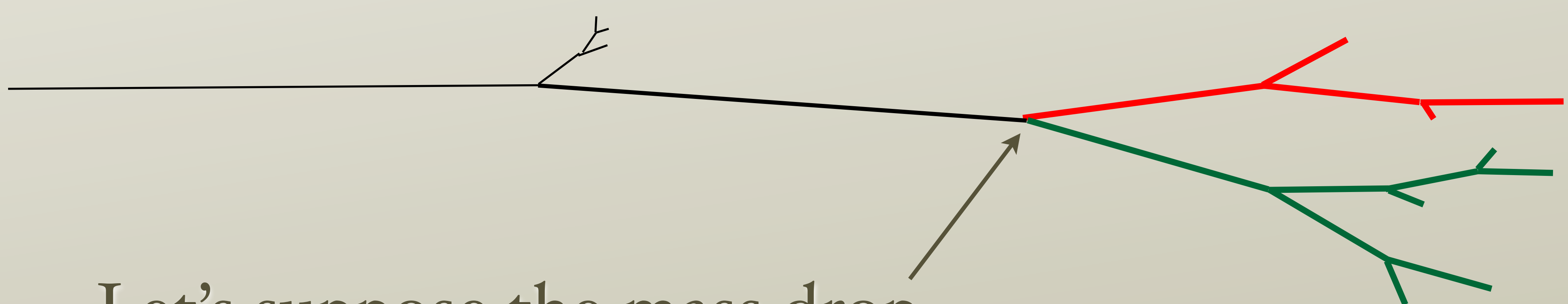
$$\max(M_i, M_j) < 0.67\, M_{\{i,j\}}$$

$$\min(p_{T,i}^2, p_{T,j}^2)[(y_i - y_j)^2 + (\phi_i - \phi_j)^2] > 0.09\, M_{\{i,j\}}^2$$



- If mass drop condition isn't met, drop smaller p_T daughter and keep looking.

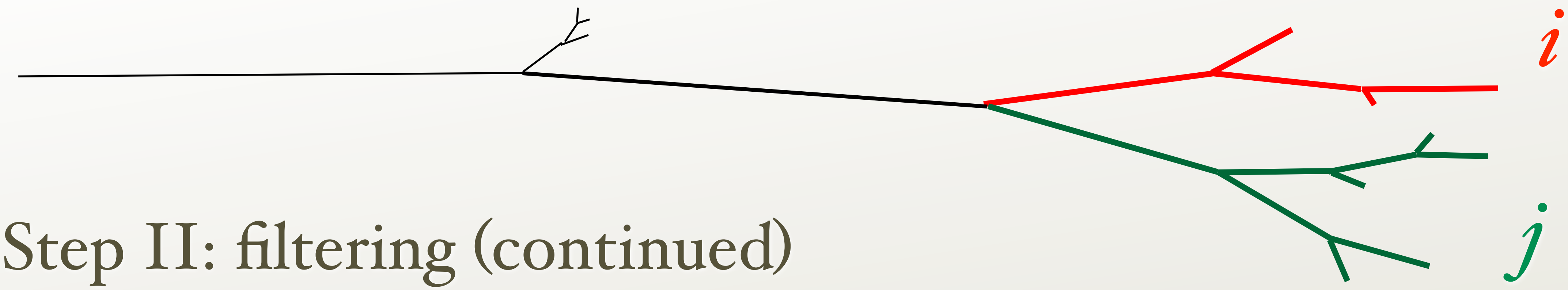
- If it is never met, remove the event from your sample.



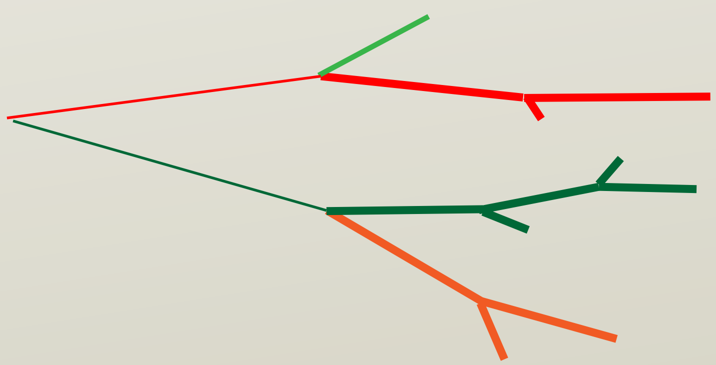
Let's suppose the mass drop condition is met here.

b-quark jets?

- Step II: filtering the prospective *b*-jets, *i* and *j*.
  - Are both prospective *b*-jets tagged as containing *b*-quarks?
  - (In simulating this, we assume a *b*-tagging efficiency of 60% and a misstag probability of 2%)
  - If i and j are not *b*-tagged, reject the event.

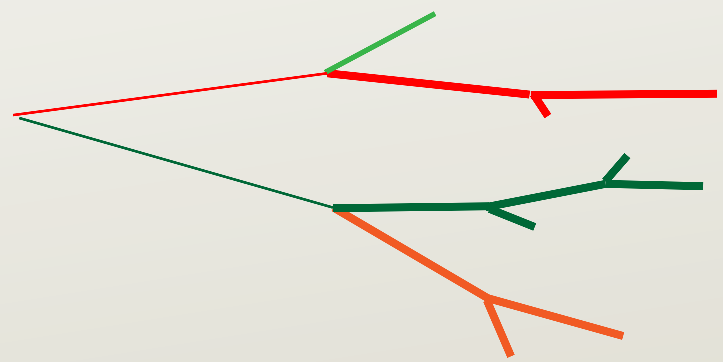


- Step II: filtering (continued)
  - Apply the C-A algorithm with to protojets i and j with
    $$R = \min\left(\frac{1}{2}[(y_i - y_j)^2 + (\phi_i - \phi_j)^2]^{1/2}, 0.3\right)$$



  - Are the two highest p_T subjets thus found tagged as containing b-quarks?
  - If not, throw out the event.

- Step II: filtering (continued some more)
  - Throw out all but the three highest p_T subjets thus found.
  - What remains is the filtered jet.

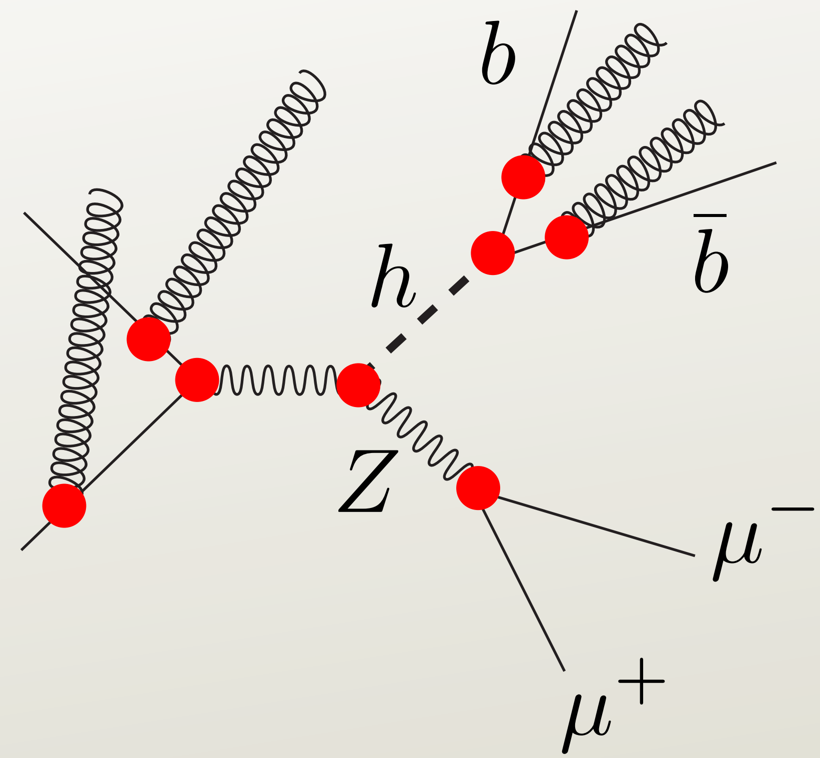


- Measure the mass of this filtered jet.

- Count event if it is in mass window $|m_{b\bar{b}} - m_H| < \Delta m_H$.
- *e.g.* $\Delta m_H = 10$ GeV

Shower deconstruction

Event selection



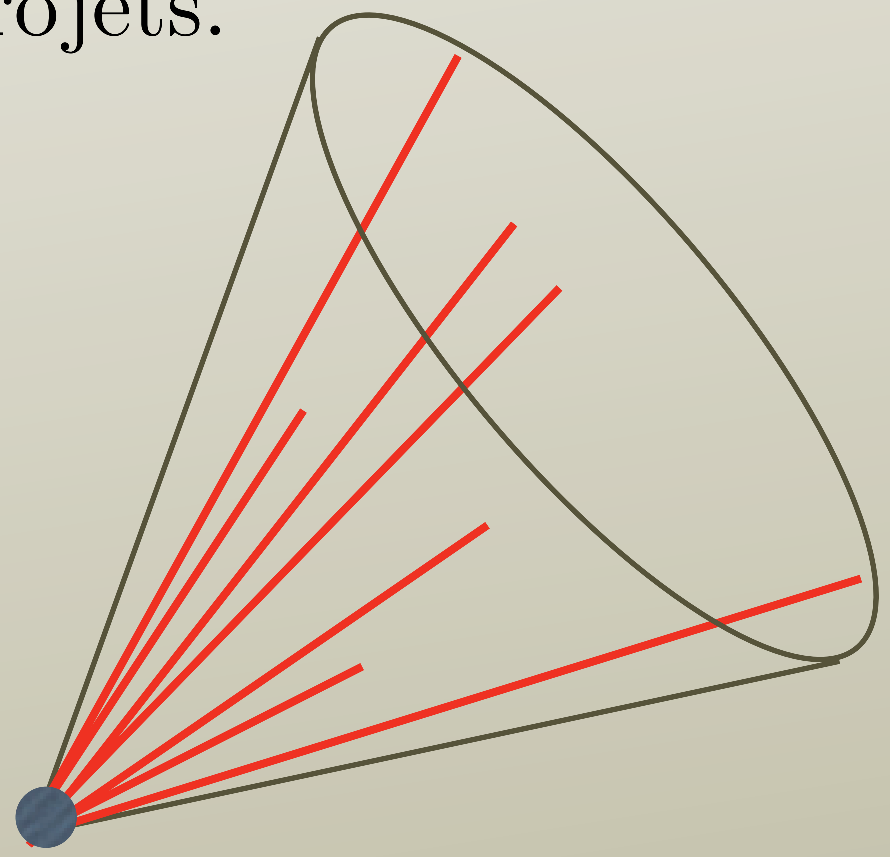
- Demand $\mu^+\mu^-$ or e^+e^- with

$$|m_{l^+l^-} - m_Z| < 10 \text{ GeV}$$
$$p_{T,l^+l^-} > 200 \text{ GeV}$$

- Combine final state hadrons in cells of size 0.1×0.1.
- Adjust $|\vec{p}|$ to make each cell momentum massless.
- Remove cells with energy less than 0.5 GeV.
- Apply anti-k_T jet algorithm with $R = 1.2$.
- The jet with the highest p_T is the "fat jet."
- Demand $p_T^{\text{fat jet}} > 200$ GeV.

Define the microjet constituents

- Use the k_T algorithm to group the fat jet into subjets.
- Use $R = 0.15$.
- This is more or less like an Atlas "topocluster."
- If too many subjets ($e.g. > 7$) drop those with smallest p_T.
- We call the resulting subjets the "microjets."
- Add 0.1 GeV to the energy of each microjet.



The variables

- Microjets described by momenta $\{p\}_N = \{p_1, \ldots, p_N\}$.
- Also provide b-tags, t_j: T, F, or "none."
  - T or F tags to three highest p_T microjets if $p_T > 15$ GeV.
  - If any hadron in microjet j contains a b or $\bar{b}$ quark,

    $t_j =$ T with a probability 0.6

    $t_j =$ F with a probability 0.4.
  - If no hadron in microjet j contains a b or $\bar{b}$ quark,

    $t_j =$ T with a probability 0.02

    $t_j =$ F with a probability 0.98

- So microjets described by momenta $\{p, t\}_N$.

What we would like

- Our data: momenta p and b-tags for N microjets, $\{p, t\}_N$.

- Define probabilities for signal and background events to have $\{p, t\}_N$ according to a trusted Monte Carlo:

$$P_{\rm MC}(\{p,t\}_N|{\rm S}) = \frac{1}{\sigma_{\rm MC}({\rm S})}\,\frac{d\sigma_{\rm MC}({\rm S})}{d\{p,t\}_N}$$

$$P_{\rm MC}(\{p,t\}_N|{\rm B}) = \frac{1}{\sigma_{\rm MC}({\rm B})}\,\frac{d\sigma_{\rm MC}({\rm B})}{d\{p,t\}_N}$$

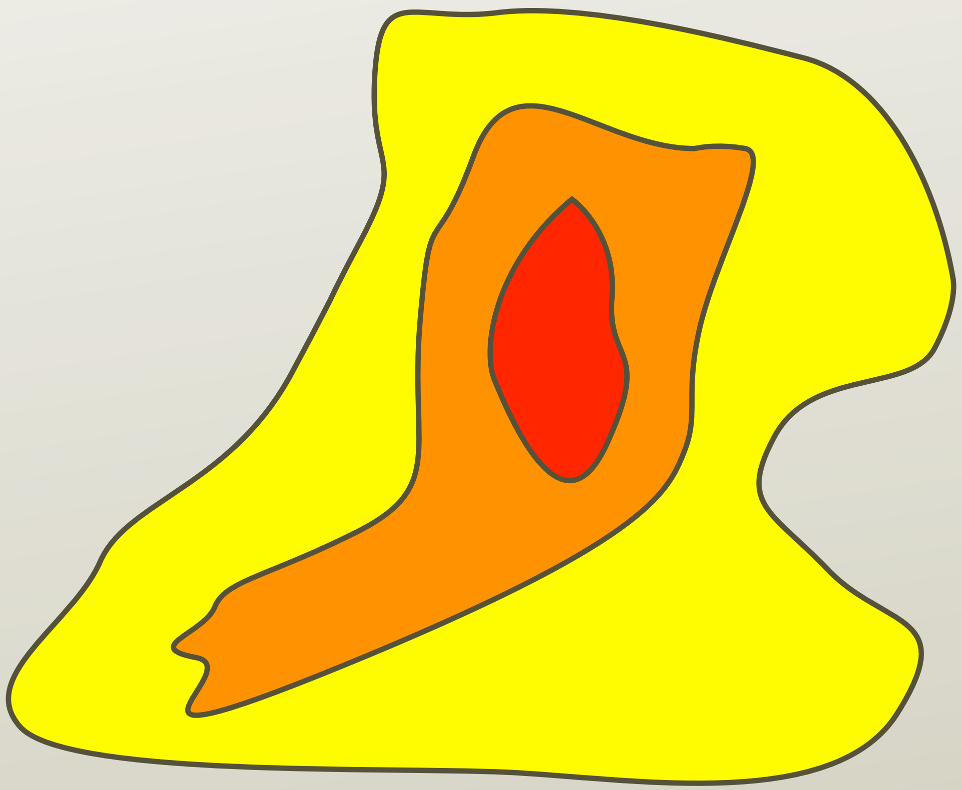
- We would like to separate signal from background using

$$\chi_{\rm MC}(\{p,t\}_N) = \frac{P_{\rm MC}(\{p,t\}_N|{\rm S})}{P_{\rm MC}(\{p,t\}_N|{\rm B})}$$

Why?

- Assuming that you believe your Monte Carlo, to get the most signal cross section for a given background cross section by making a cut, your cut should be along a contour line of

$$\chi_{\rm MC}(\{p,t\}_N) = \frac{P_{\rm MC}(\{p,t\}_N|{\rm S})}{P_{\rm MC}(\{p,t\}_N|{\rm B})}$$



What we do

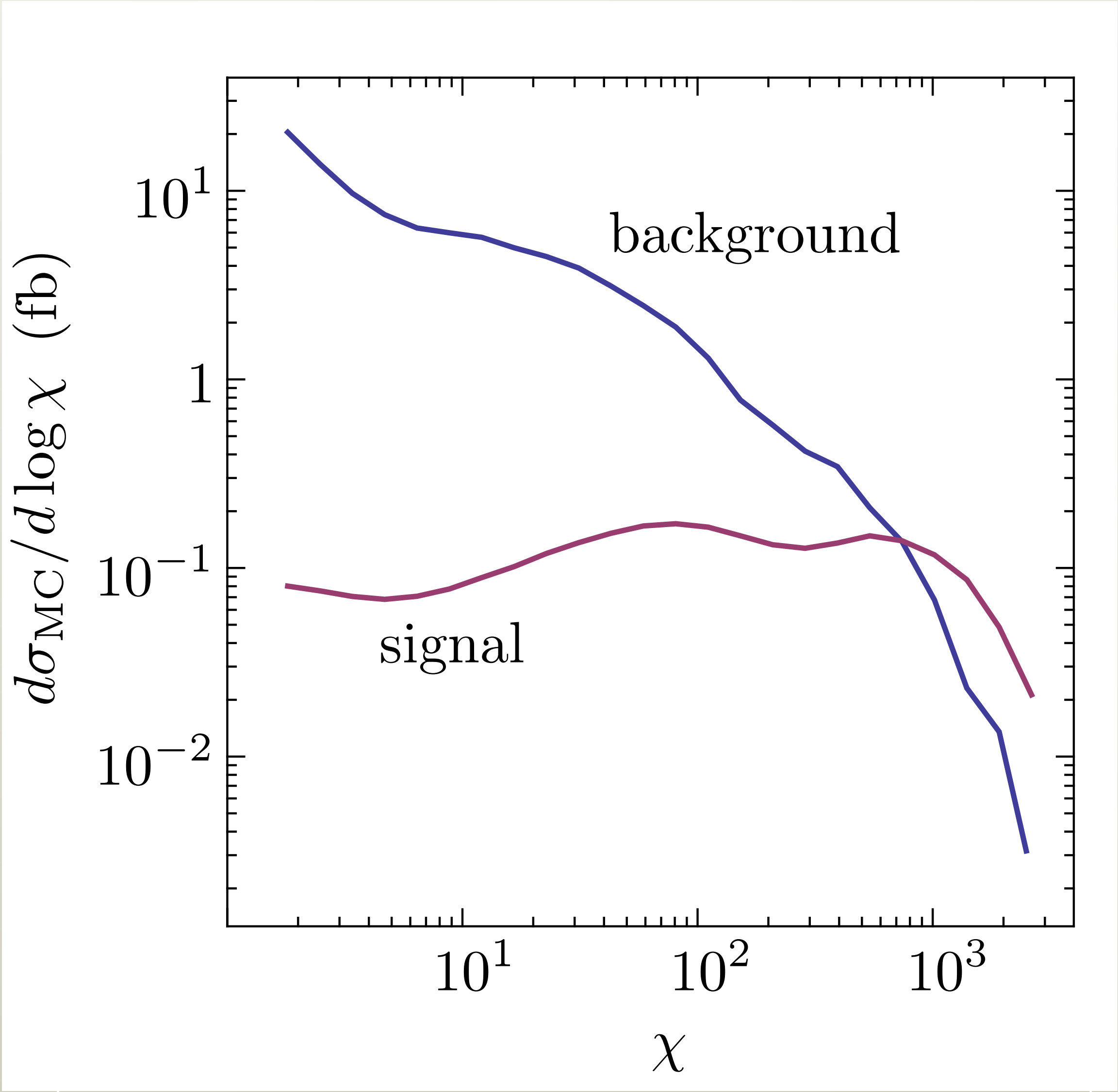
- Calculate

$$\chi(\{p,t\}_N) = \frac{P(\{p,t\}_N|\mathrm{S})}{P(\{p,t\}_N|\mathrm{B})}$$

according to a "simplified parton shower" algorithm.

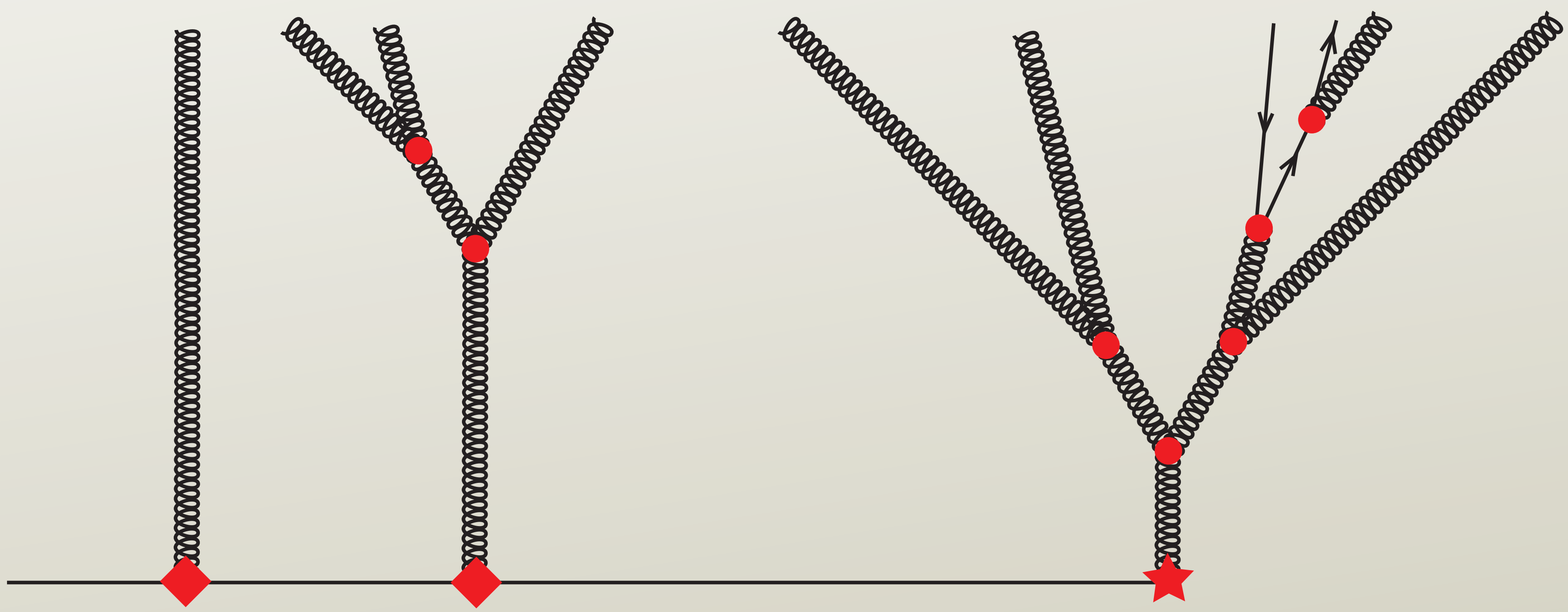
Result

- We calculate χ for samples of signal and background events generated with PYTHIA.



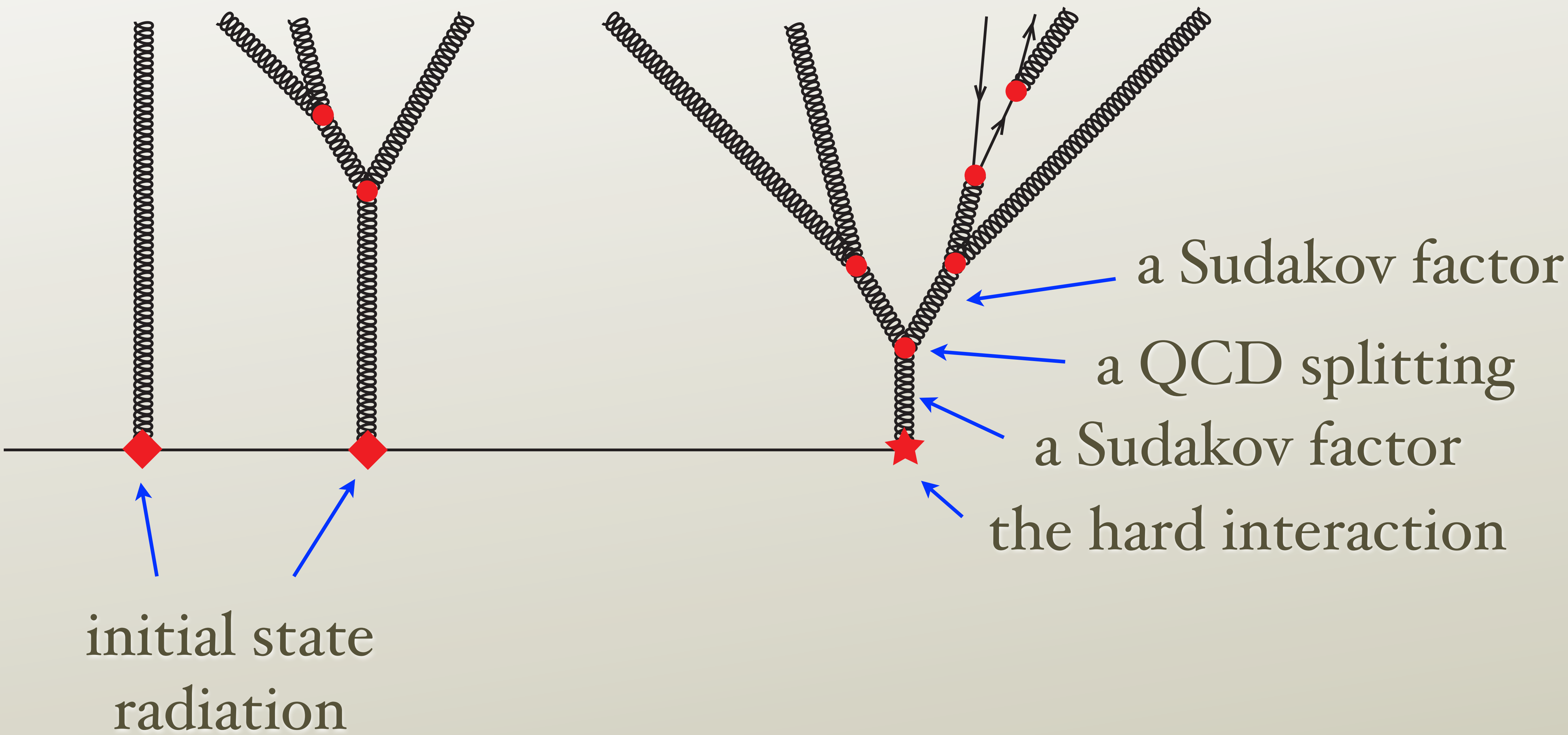
How does it work?

- We sum over event histories.



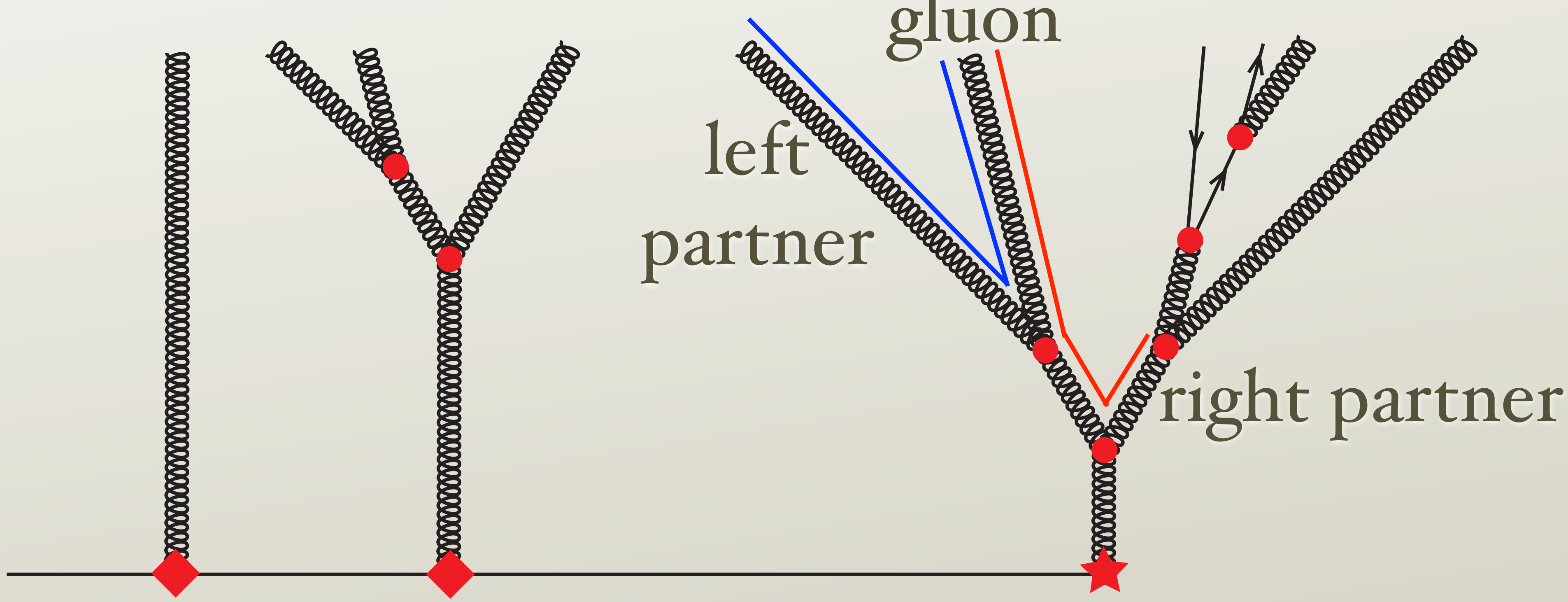
- Each vertex and propagator corresponds to a shower algorithm factor.

About histories



Color connections

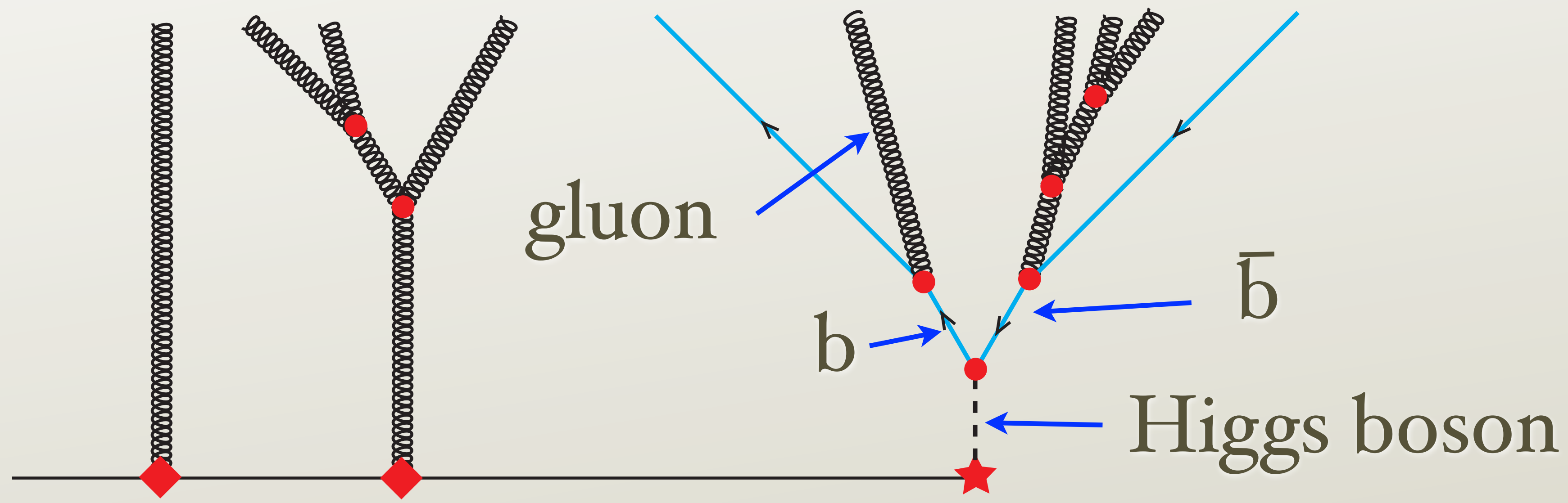
Each gluon has two color connected partners.

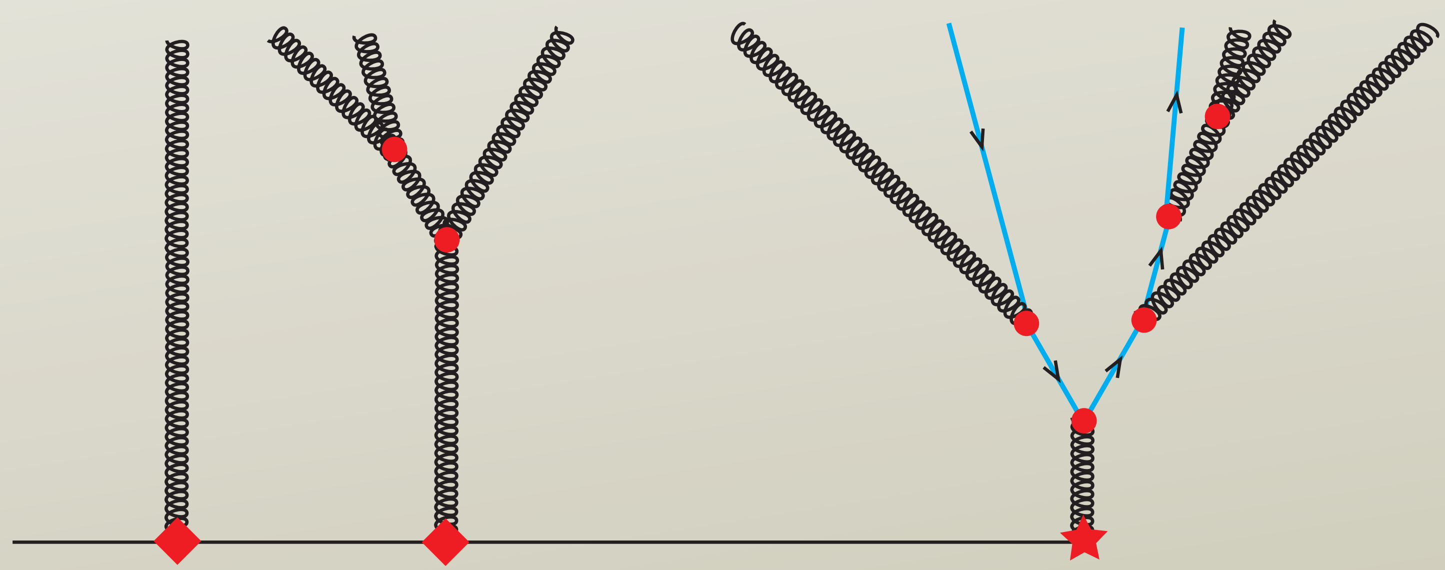


Each quark has one color connected partner.

Some partners are unknown, likely outside the fat jet.

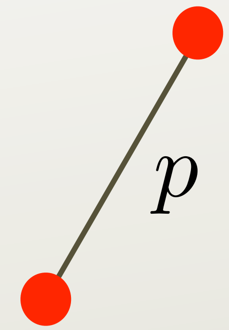
There are more kinds of histories



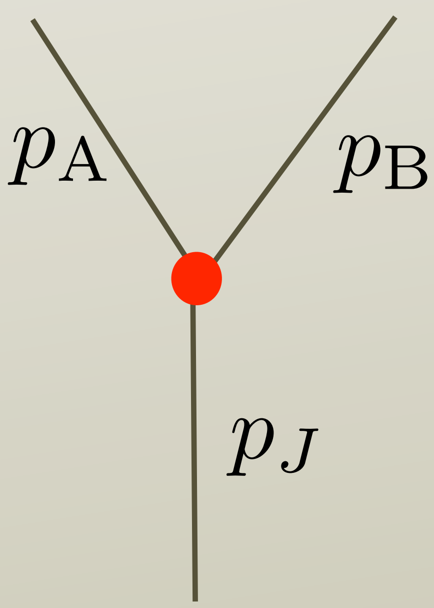


background process with $g \rightarrow b\,\bar{b}$ splitting

Kinematics



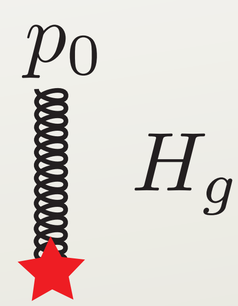
$$p = \left(\frac{1}{\sqrt{2}} \sqrt{k^2 + \mu^2}\, e^{y}, \frac{1}{\sqrt{2}} \sqrt{k^2 + \mu^2}\, e^{-y}, k \cos\phi, k \sin\phi \right)$$

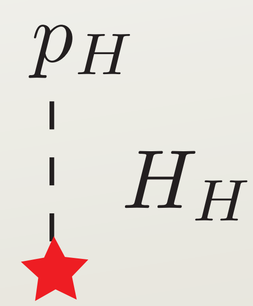


$$p_J = p_{\mathrm{A}} + p_{\mathrm{B}}$$

The daughter partons are not on shell.

The hard interaction



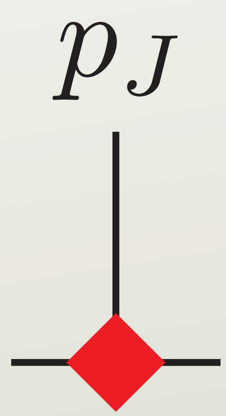


$$H_g = N_{\rm pdf} \left(\frac{p_{T,\min}^2}{k_0^2} \right)^{N_{\rm pdf}} \frac{1}{k_0^2} \; \Theta(k_{T,\rm I}^2 < Q^2/4)$$

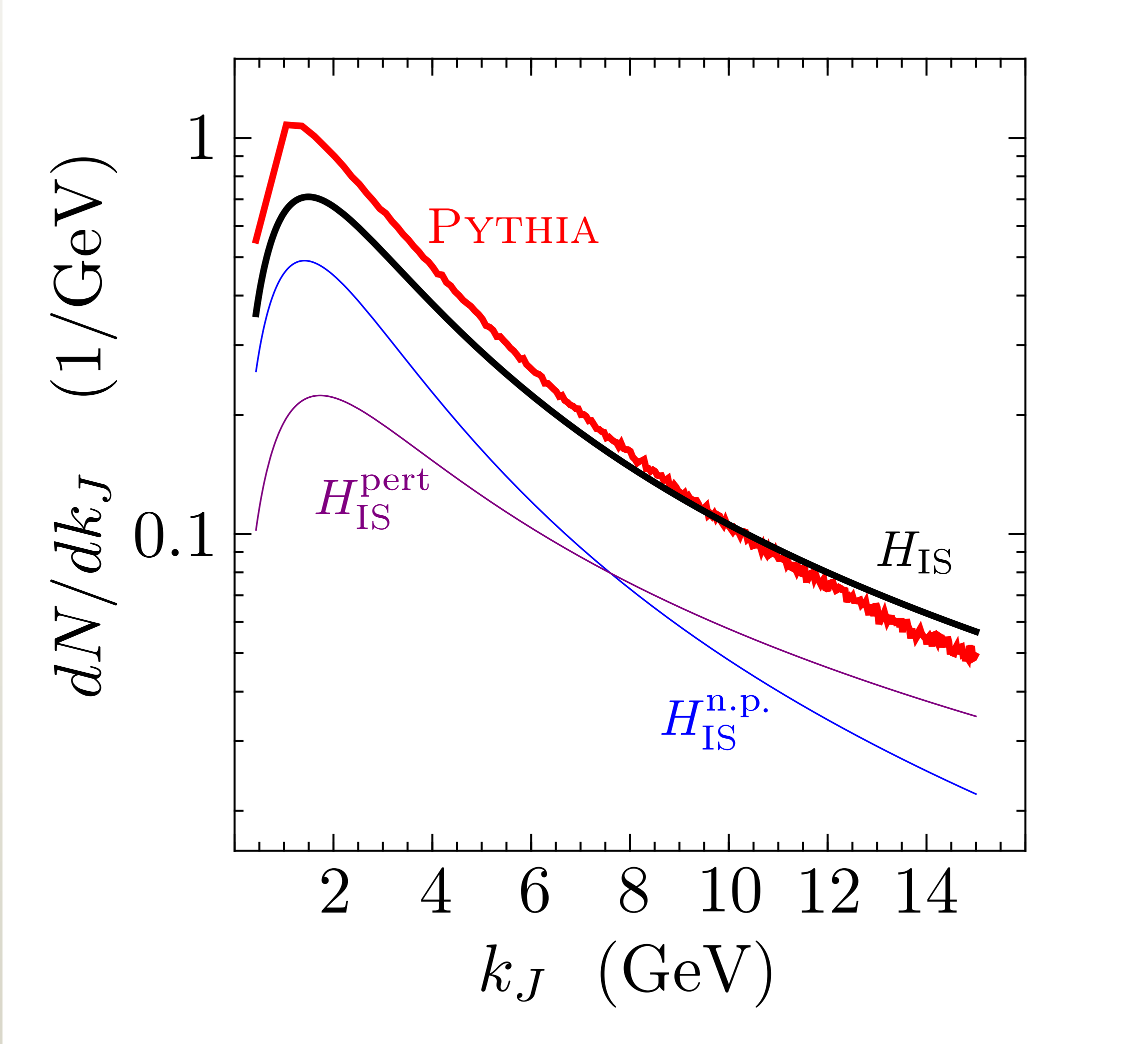
$$H_H = N_{\rm pdf} \left(\frac{p_{T,\min}^2 + m_H^2}{k_H^2 + m_H^2} \right)^{N_{\rm pdf}} \frac{1}{k_H^2 + m_H^2} \; \Theta(k_{T,\rm I}^2 < Q^2/4)$$

$$Q^2 = \left(\sum_{i \in \text{fat jet}} \vec{p}_{T,i} \right)^2 + \left(\sum_{i \in \text{fat jet}} p_i \right)^2$$

Initial state radiation

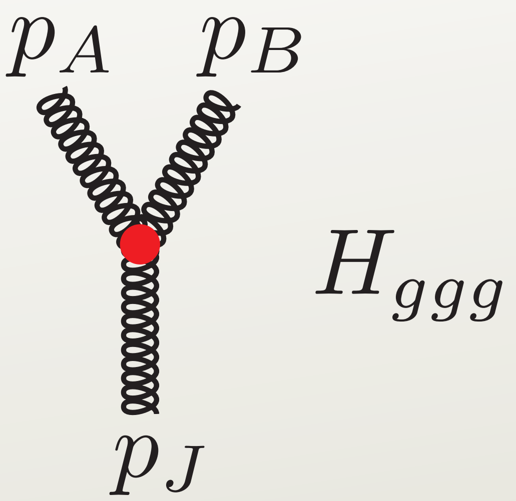


H_{IS}



$$H_{\mathrm{IS}} = \frac{\alpha_s(k_J^2+\kappa_{\mathrm{p}}^2)}{k_J^2+\kappa_{\mathrm{p}}^2}\,\frac{8\pi\,C_{\mathrm{A}}}{(1+c_R\,k_J/Q)^{n_R}} + \frac{16\pi\,c_{\mathrm{np}}(\kappa_{\mathrm{np}}^2)^{n_{\mathrm{np}}-1}}{[k_J^2+\kappa_{\mathrm{np}}^2]^{n_{\mathrm{np}}}}$$

Gluon splitting



"s" $=$ softer of A and B

"h" $=$ harder of A and B

$$\frac{[1 - z(1-z)]^2}{z(1-z)}$$

angle factor

$$H_{ggg} = 8\pi C_A \, \frac{\alpha_s(\mu_J^2)}{\mu_J^2} \, \frac{k_J^2}{k_s k_h} \left[1 - \frac{k_s k_h}{k_J^2}\right]^2 \frac{\theta_{hk}^2}{\theta_{sh}^2 + \theta_{sk}^2}$$

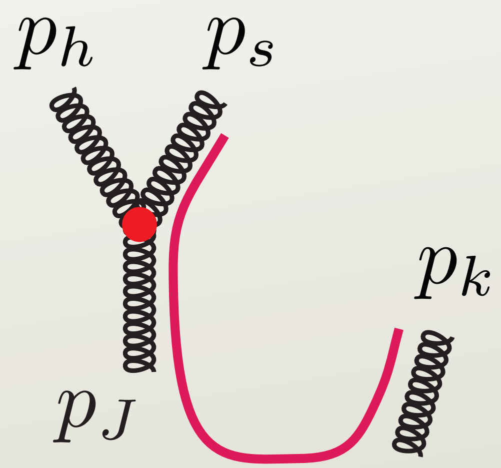
$$\times \Theta\left(2\,\frac{\mu_J^2}{k_J} < \frac{\mu_K^2}{k_K}\right)$$

$\mu_J^2 = p_J^2$

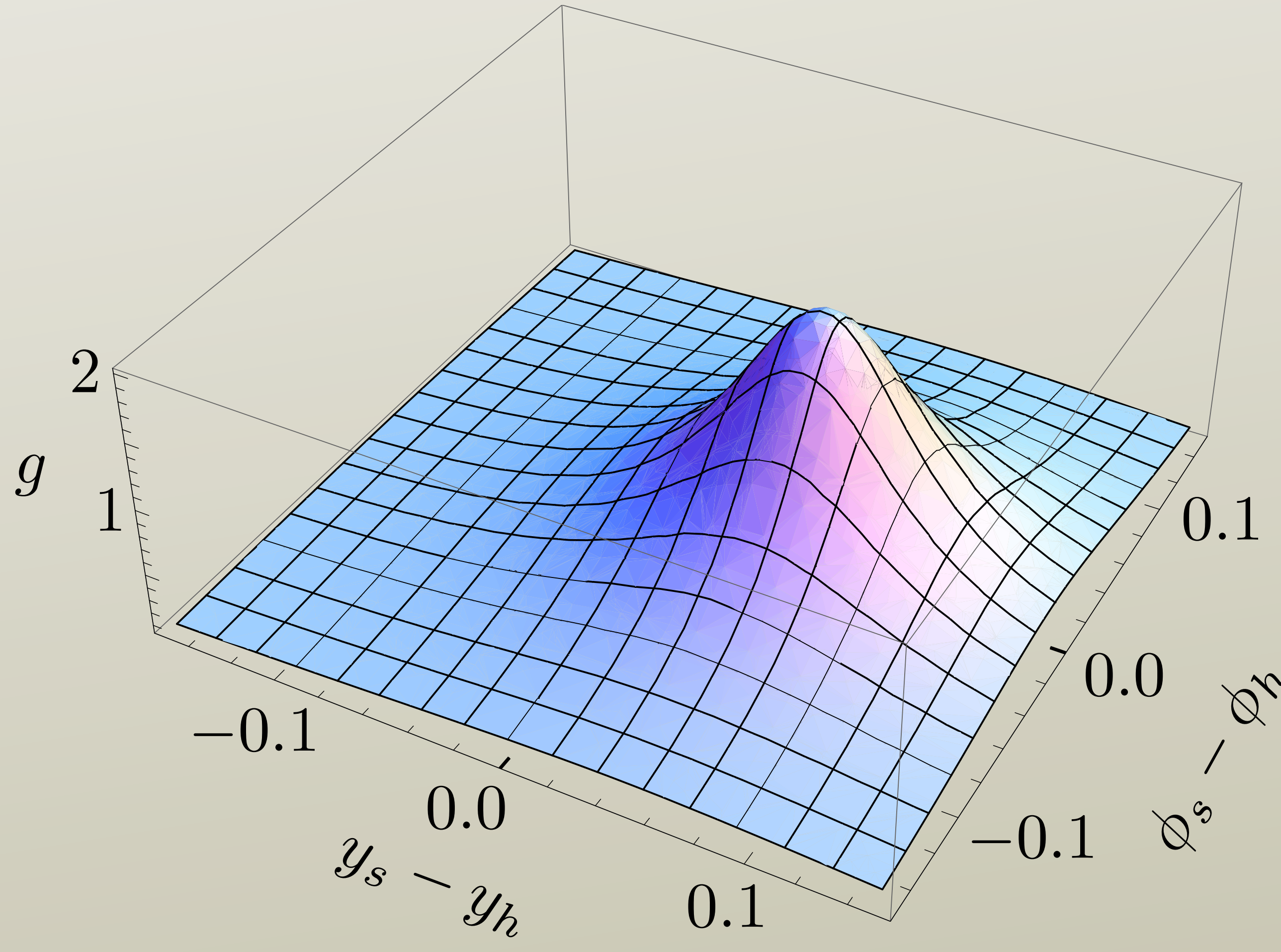
ordering of shower times

"K" $=$ grandmother parton

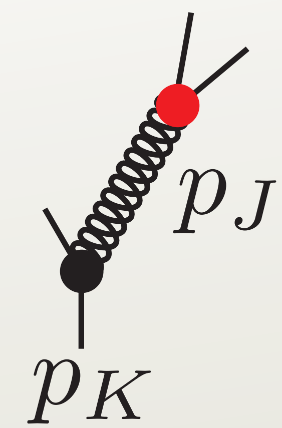
The angle factor



$$g = \frac{\theta_{hk}^2}{\theta_{sh}^2 + \theta_{sk}^2}$$



Sudakov factor



$$e^{-S_g}$$

$$S_g = S_{ggg}\,\Theta(S_{ggg} > 0) + n_{\mathrm{f}} S_{g\bar{q}q}$$

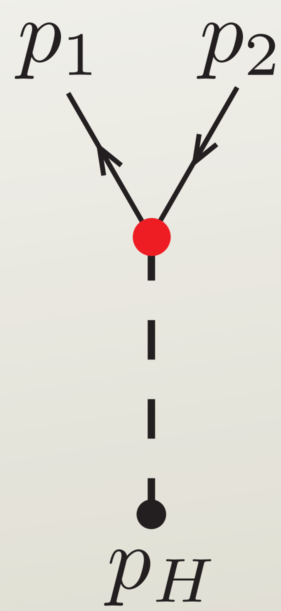
$$S_{ggg} \approx \int \frac{d\bar{\mu}_J^2}{\bar{\mu}_J^2}\;\Theta\!\left(\mu_J^2 < \bar{\mu}_J^2 < \frac{k_J}{2k_K}\,\mu_K^2\right)\;\frac{\alpha_{\mathrm{s}}(\bar{\mu}_J^2)}{2\pi}$$

$$\times \int dz\;\Theta\!\left(\frac{1}{\theta_{k(A)}^2}\,\frac{\bar{\mu}_J^2}{k_J^2} < z < 1 - \frac{1}{\theta_{k(B)}^2}\,\frac{\bar{\mu}_J^2}{k_J^2}\right)$$

$$\times C_{\mathrm{A}}\,\frac{[1 - z(1-z)]^2}{z(1-z)}$$

limits on z
from the angle factor
(approximately)

Higgs decay



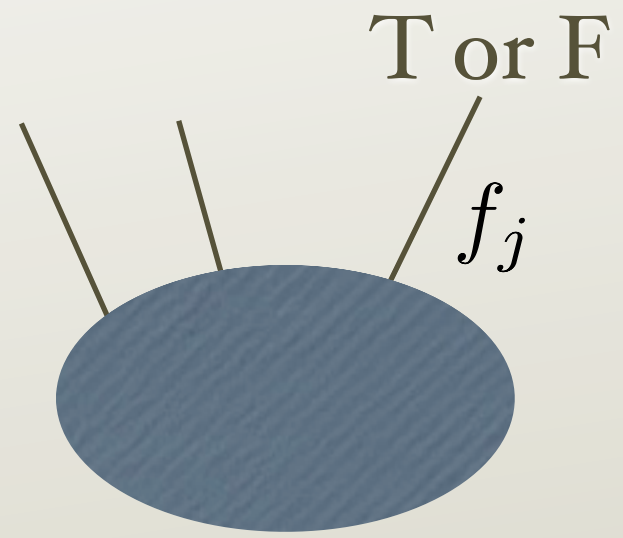
$$H_{Hb\bar{b}}\, \exp(-S_{Hb\bar{b}})$$

$$He^{-S} = 16\pi^2 \, \frac{\Theta(|m_{b\bar{b}} - m_H| < \Delta m_H)}{4 m_H \, \Delta m_H}$$

default: $\Delta m_H = 10 \text{ GeV}$

Probabilities for *b*-tags

- b tags are assigned to the three highest p_T microjets (with $p_T > 15$ GeV).



- If microjet j is a b or $\bar{b}$, we say that $t_j = T$ with probability 0.6 and $T_j = F$ with probability 0.4.

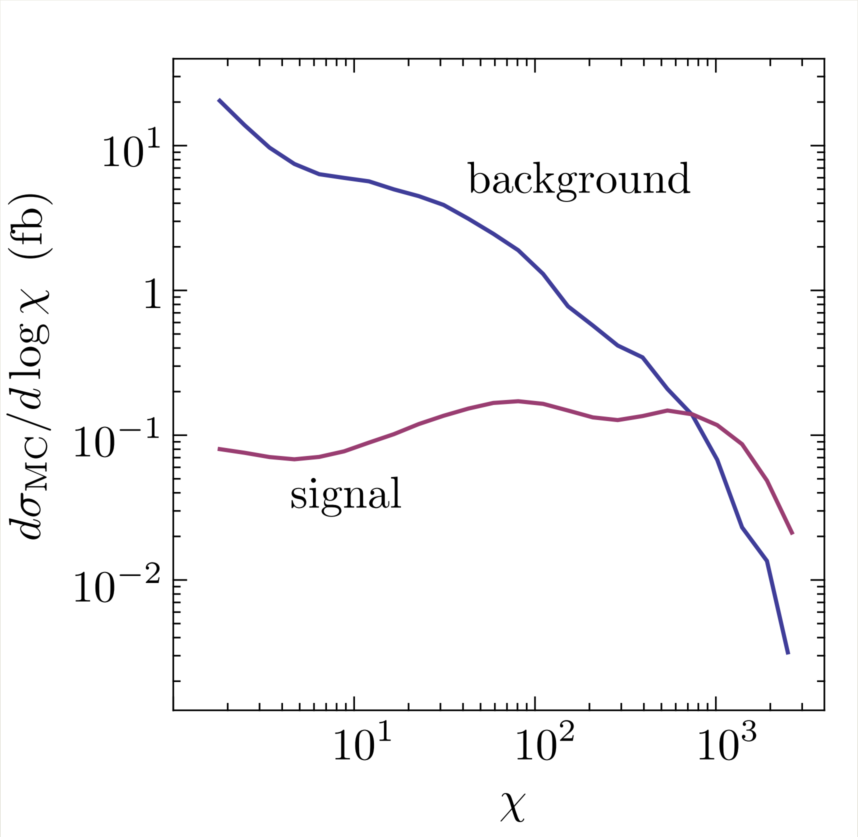
- If microjet j is not a b or $\bar{b}$, we say that $t_j = T$ with probability 0.02 and $T_j = F$ with probability 0.98.

Results

- Best to construct likelihood ratio, but let's use a simple cut.
- Define signal and background cross sections above a cut:

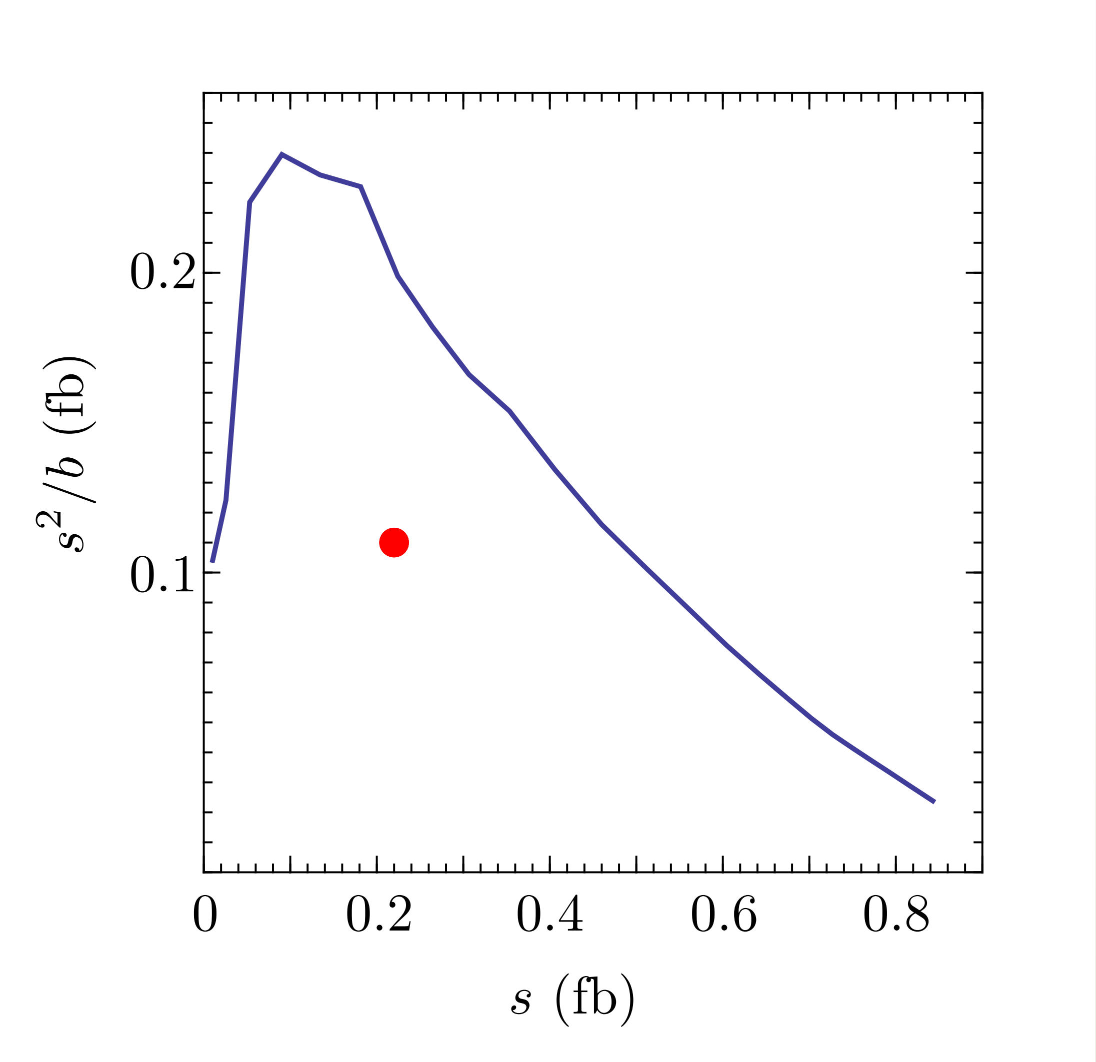
$$s(\chi) = \int_\chi^\infty d\bar{\chi}\, \frac{d\sigma_{\rm MC}(\mathrm{S})}{d\bar{\chi}}$$

$$b(\chi) = \int_\chi^\infty d\bar{\chi}\, \frac{d\sigma_{\rm MC}(\mathrm{B})}{d\bar{\chi}}$$



- We can choose the signal cross section s by adjusting χ.
- We try to make the statistical significance s^2/b large.

- We find



- *e.g.* with $\int dL = 100\ \mathrm{fb}^{-1}$ we can choose $s = 0.1$ fb.
- Then $N(S) = 10$.
- $s^2/b = 0.25$ fb gives $N(S)^2/N(B) = 25$.
- That is $N(S)/\sqrt{N(B)} = 5$.
- The red point is what you get with the BDRS method.

Conclusions

- Shower deconstruction needs a lot of development.
- So far, it is a little better than existing methods.
- It is modular and the parts can be improved.
- We expect that it will be useful for complicated problems.