A CLOSER LOOK: SEARCHING FOR THE RARE DECAY $B_s^0 \to \mu^+\mu^-$

Michelle Prewitt – Rice University

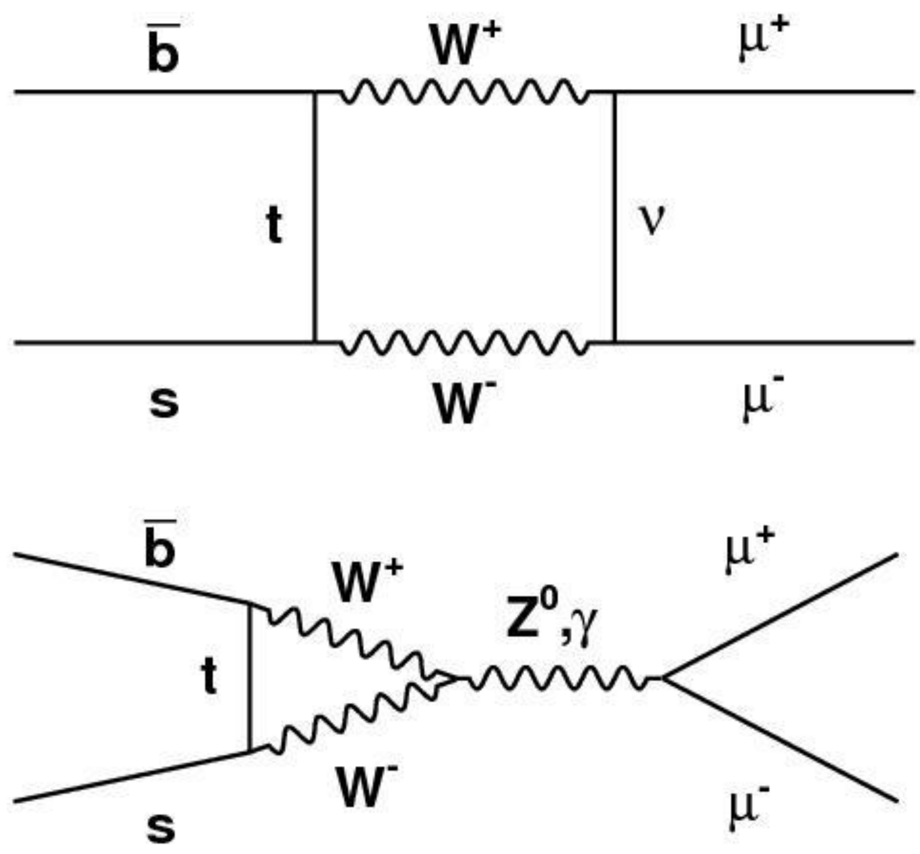
on behalf of the D0 Collaboration

February 8th, 2013

Fermilab Joint Theoretical – Experimental Seminar

Standard Model

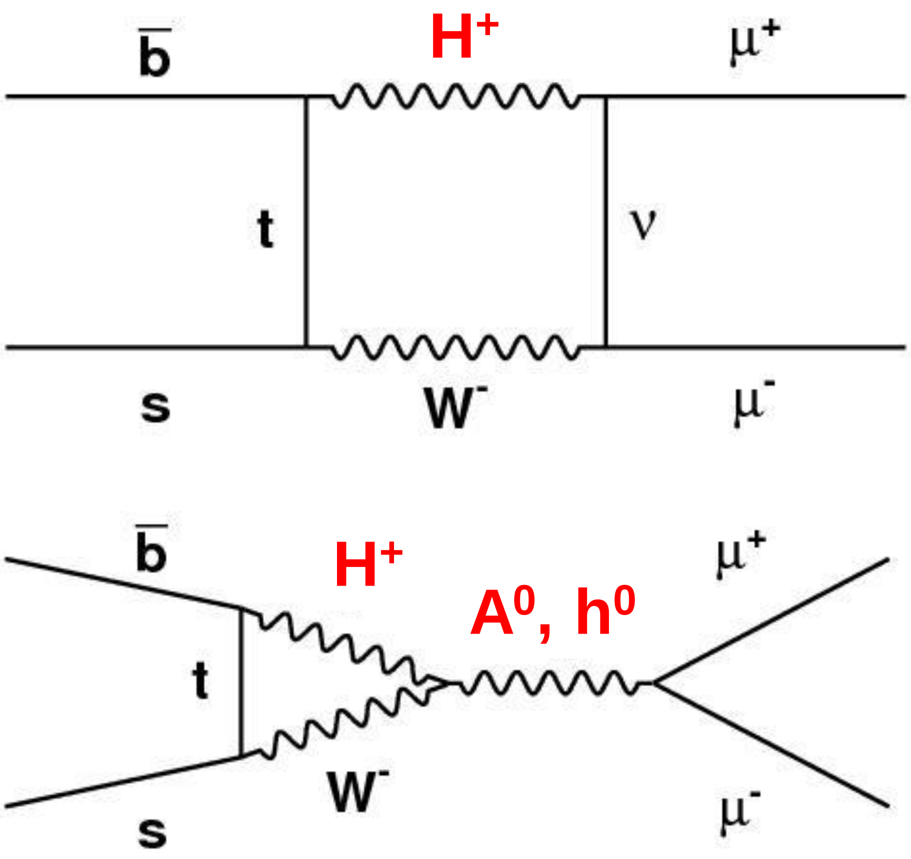
- Flavor changing neutral current restricted
- Helicity suppressed
- $BR(B^0_s \to \mu^+\mu^-)$ =
  - $(3.2 \pm 0.2) \times 10^{-9}$
  - $(3.5 \pm 0.2) \times 10^{-9}$ (lifetime corrected)*



*A. J. Buras et al., Eur. Phys. J. C72, 2172 (2012).

Standard Model

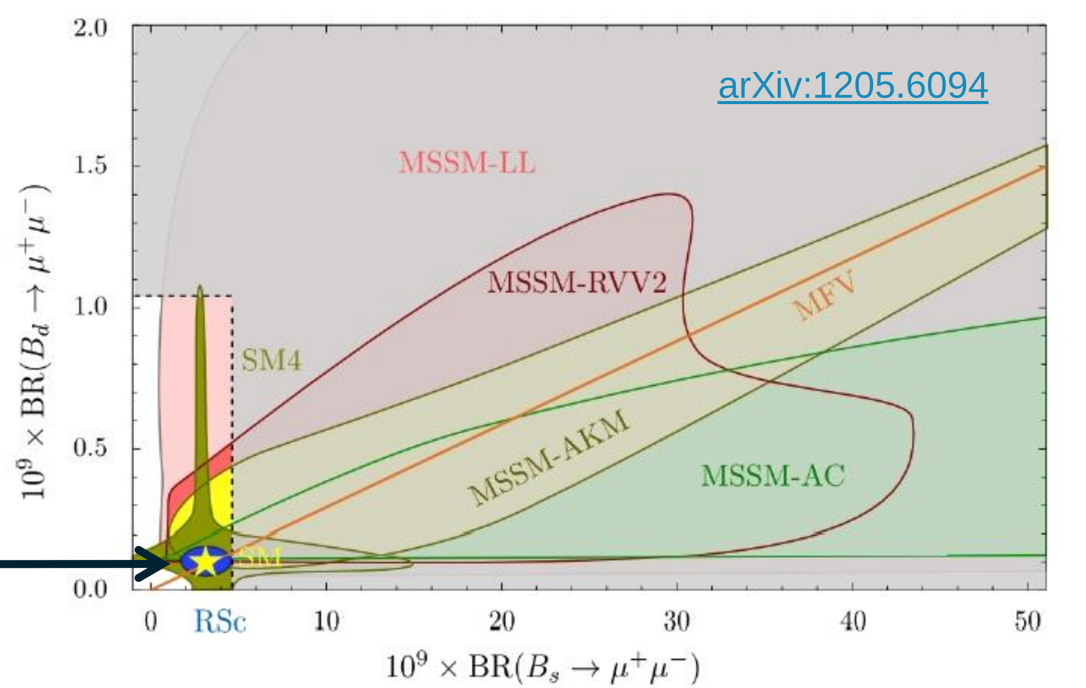
- Flavor changing neutral current restricted
- Helicity suppressed
- $BR(B^0_s \to \mu^+\mu^-) =$
  - $(3.2 \pm 0.2) \times 10^{-9}$
  - $(3.5 \pm 0.2) \times 10^{-9}$ (lifetime corrected)*



*A. J. Buras et al., Eur. Phys. J. C72, 2172 (2012).

Beyond the Standard Model

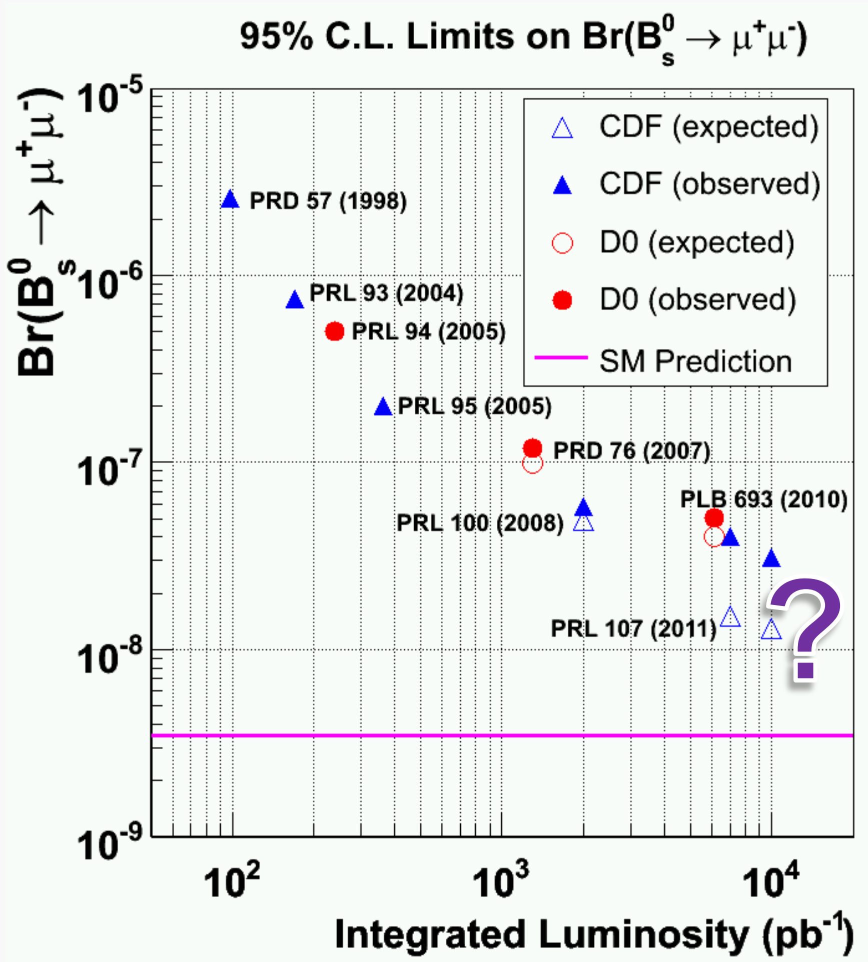
- Important decay for BSM physics because many models predict substantial enhancement or even suppression of the branching ratio



*Straub Moriond 2012

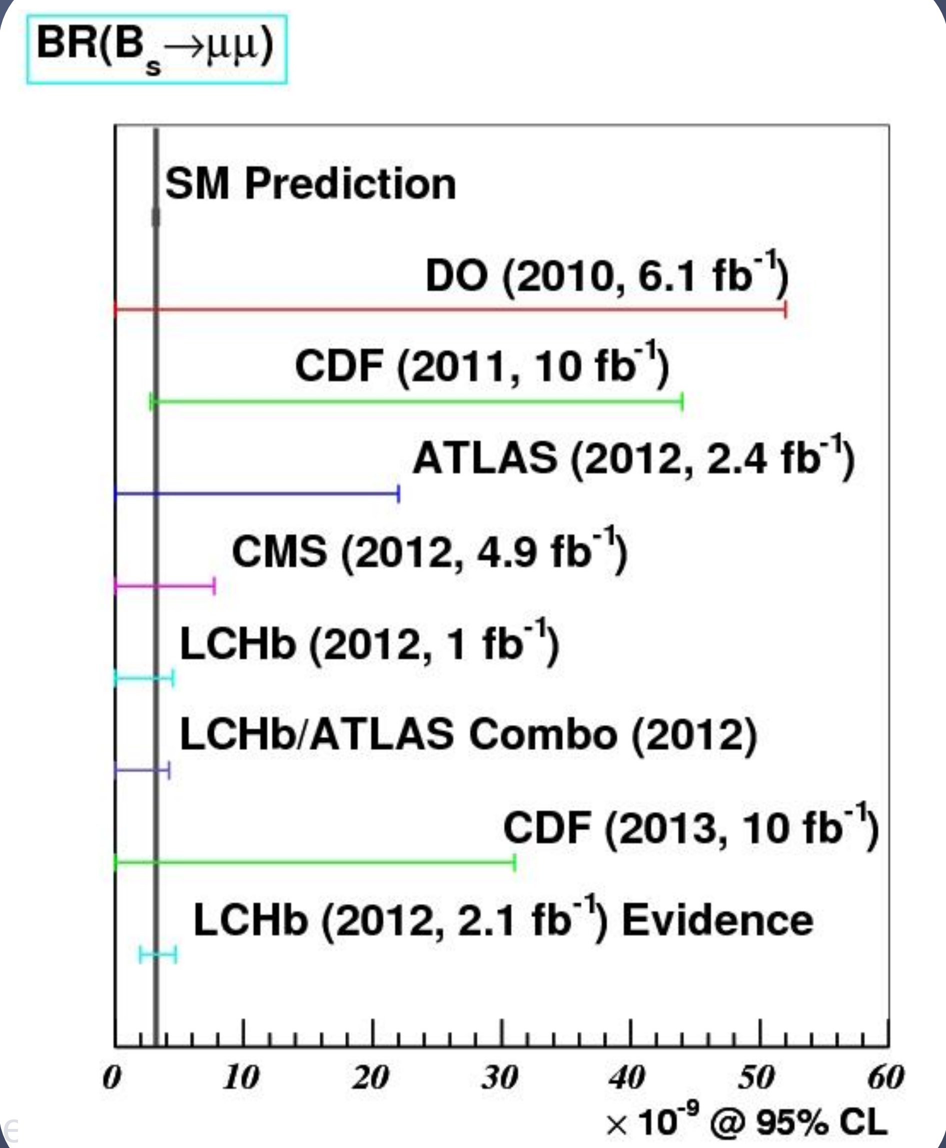
Standard Model

History of TeV Measurements



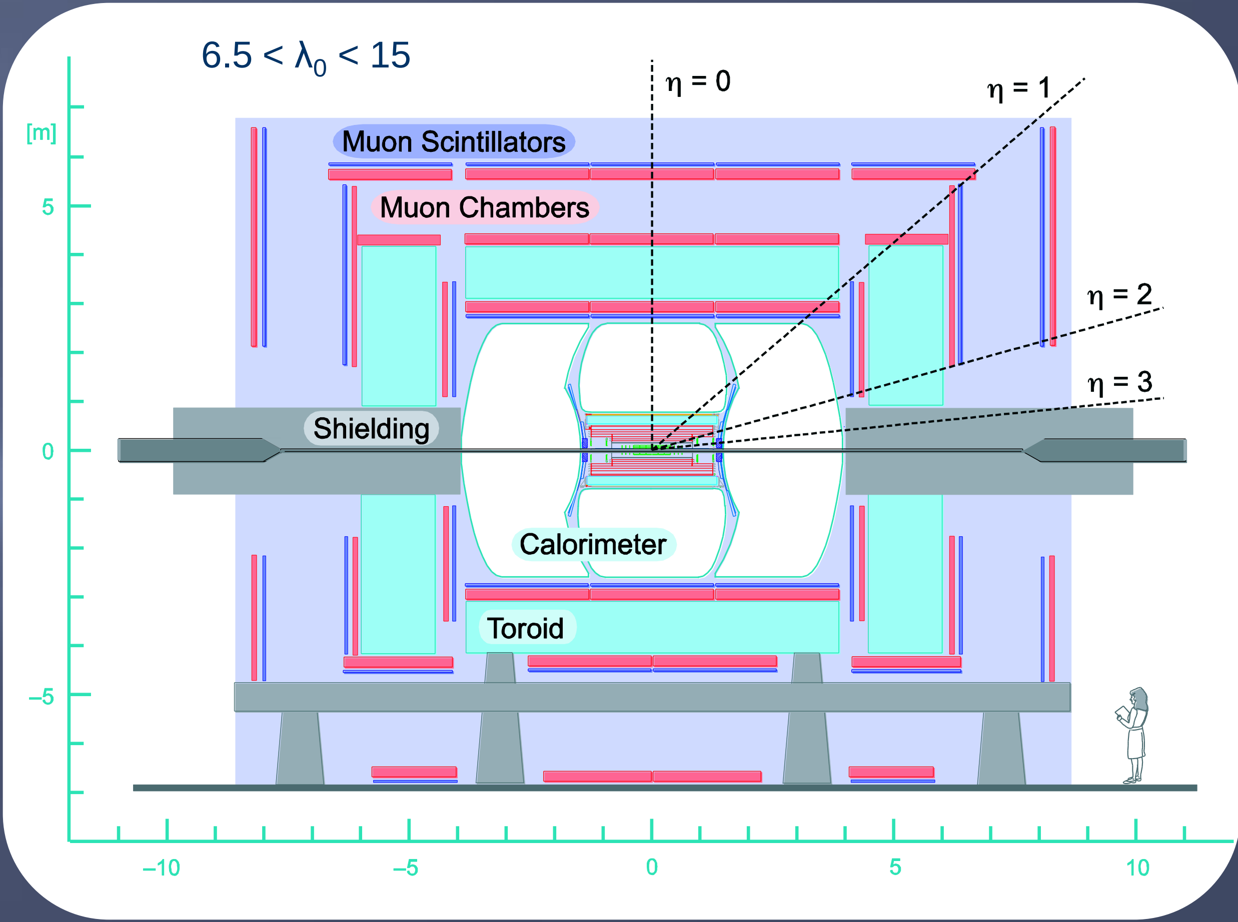
Current Experimental Situation

- LCHb:
  - Evidence
  - $BR(B^0_s \to \mu^+\mu^-) = 3.2^{+1.5}_{-1.2} \times 10^{-9}$
- CDF:
  - "…. . we measure $B(B^0_s \to \mu^+\mu^-) = 1.3^{+0.9}_{-0.7} \times 10^{-8}$ and the following bounds are set, . . . $0.8 \times 10^{-9} < B(B^0_s \to \mu^+\mu^-) < 3.4 \times 10^{-8}$ at . . . 95% C.L., respectively. "

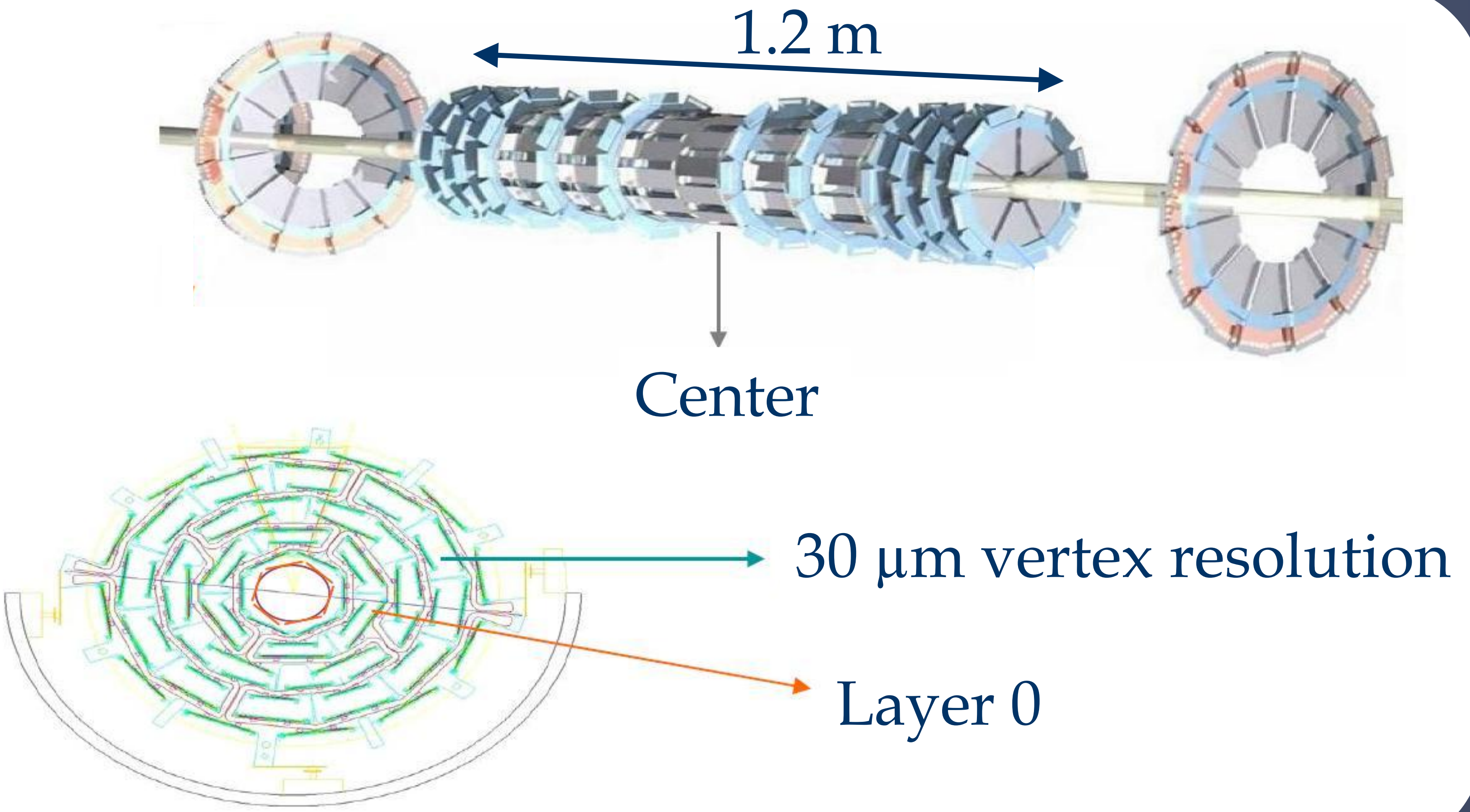


D0 Detector

- Silicon Microstrip Tracker
- Central Fiber Tracker
- Calorimeter
- Muon System
  - Coverage up to $|\eta| = 2$

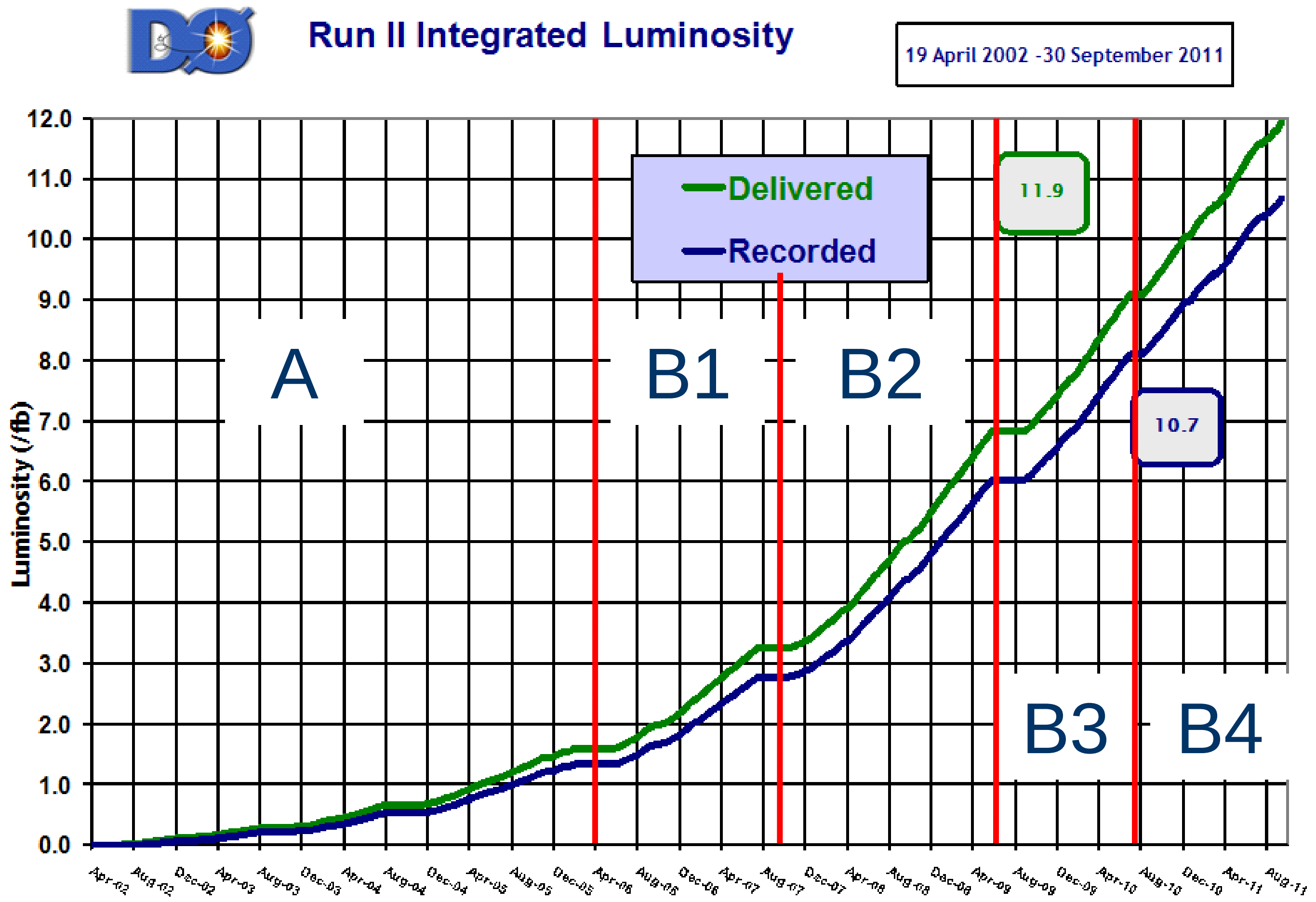


Silicon Microstrip Tracker



D0 Data

Thanks to all Operations and AD personnel!



5 data epochs in Run II

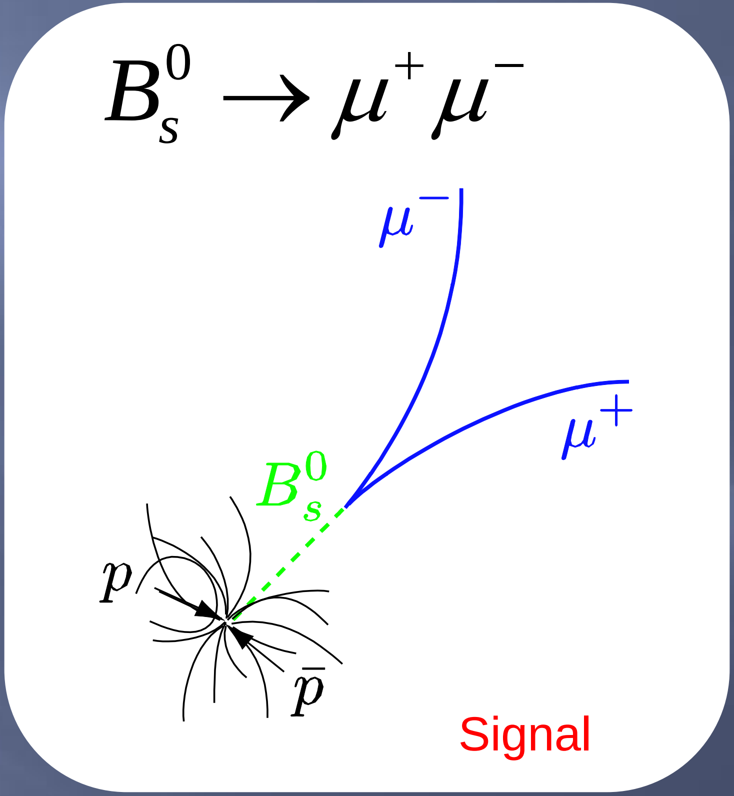
D0 Results

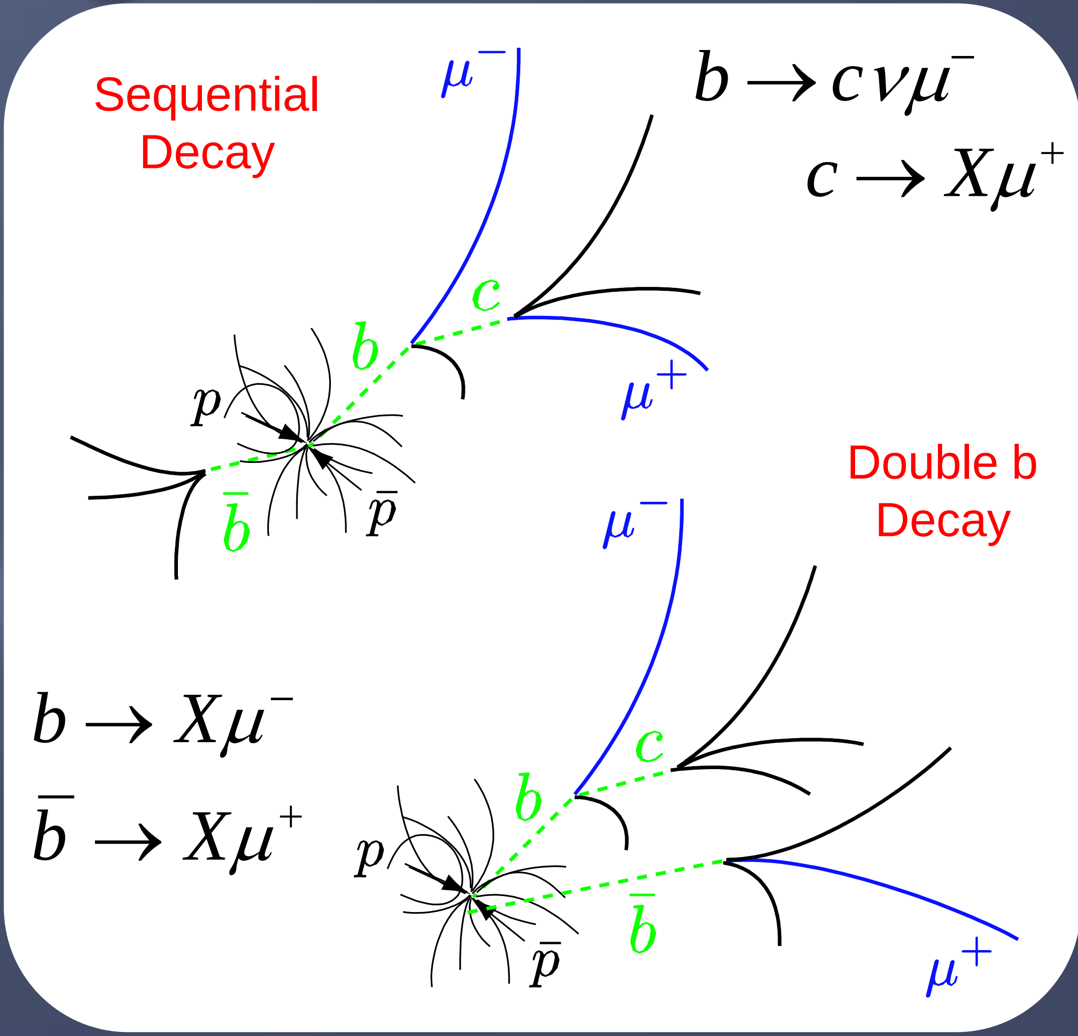
∫ Luminosity	95% BR($B_s \to \mu\mu$) <	Analysis Improvements
240 pb^{-1}	5×10^{-7}	Cut based
1.3 fb^{-1}	1.2×10^{-7}	Likelihood ratio
6.1 fb^{-1}	5.1×10^{-8}	NN, higher luminosity
10.4 fb^{-1}	???	BDT, background rejection

- New to this result:
- Full D0 data set
- Boosted Decision Tree (BDT)
- Improved background characterization

Thanks to
B. Casey
for figures!

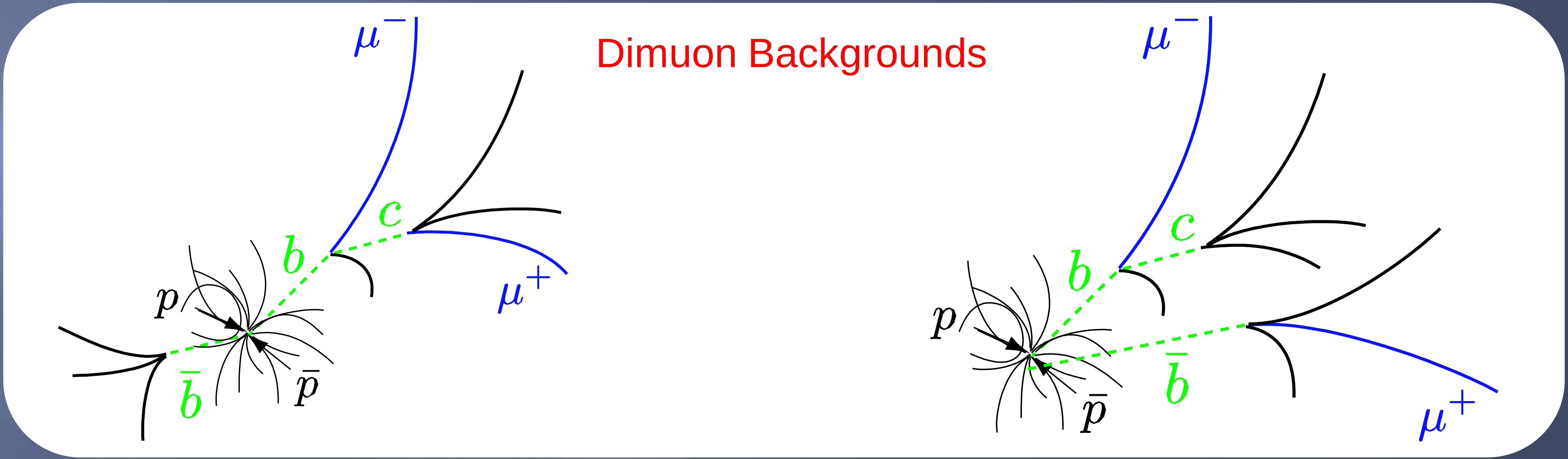
Signal vs Background



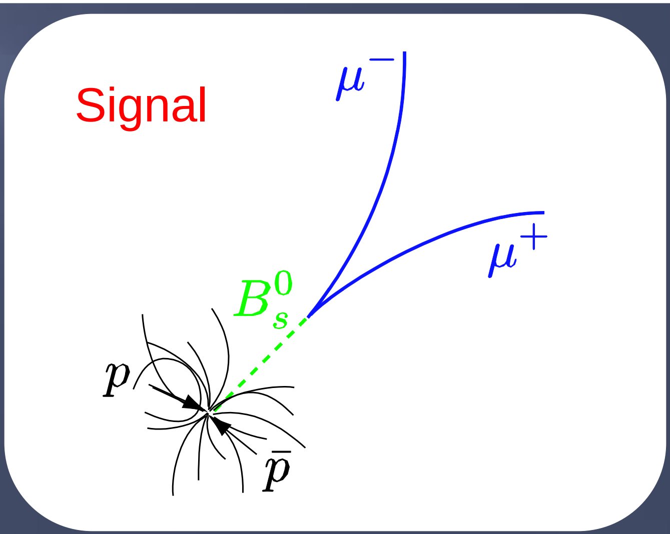


- New Variables

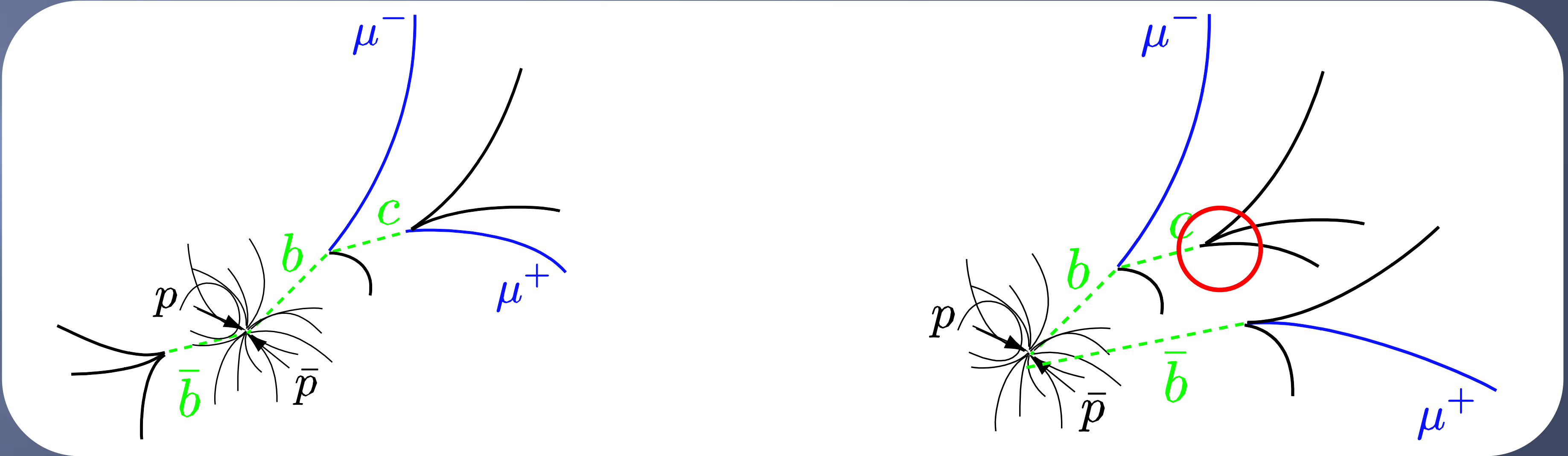
Additional Vertices



- Background has many more tracks around the dimuon vertex

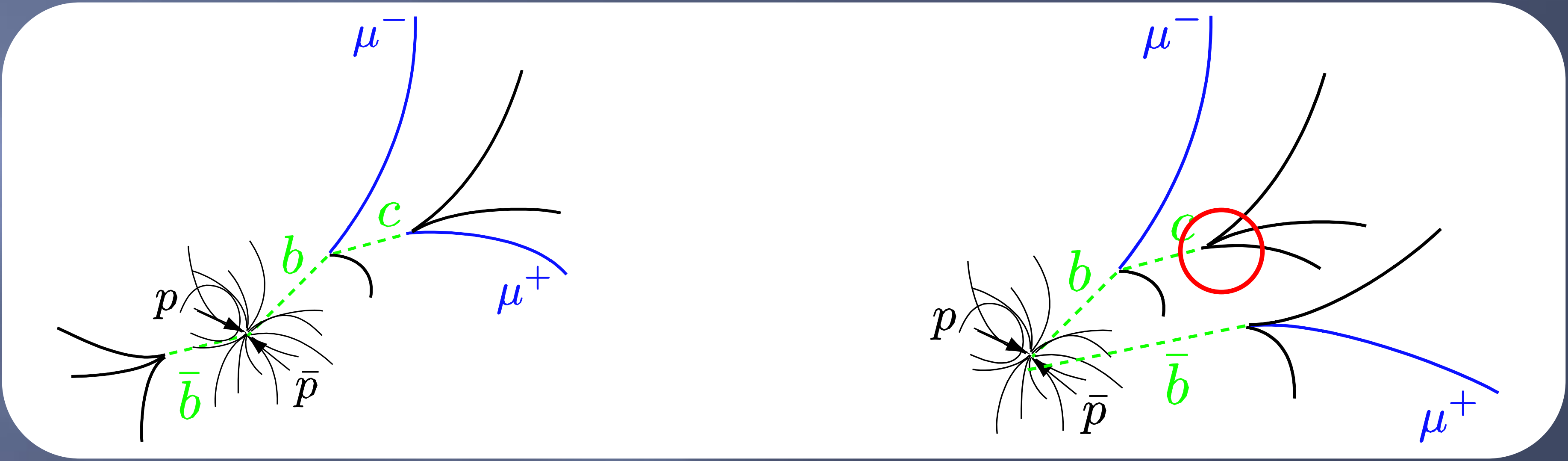


Additional Vertices



- Tertiary vertex

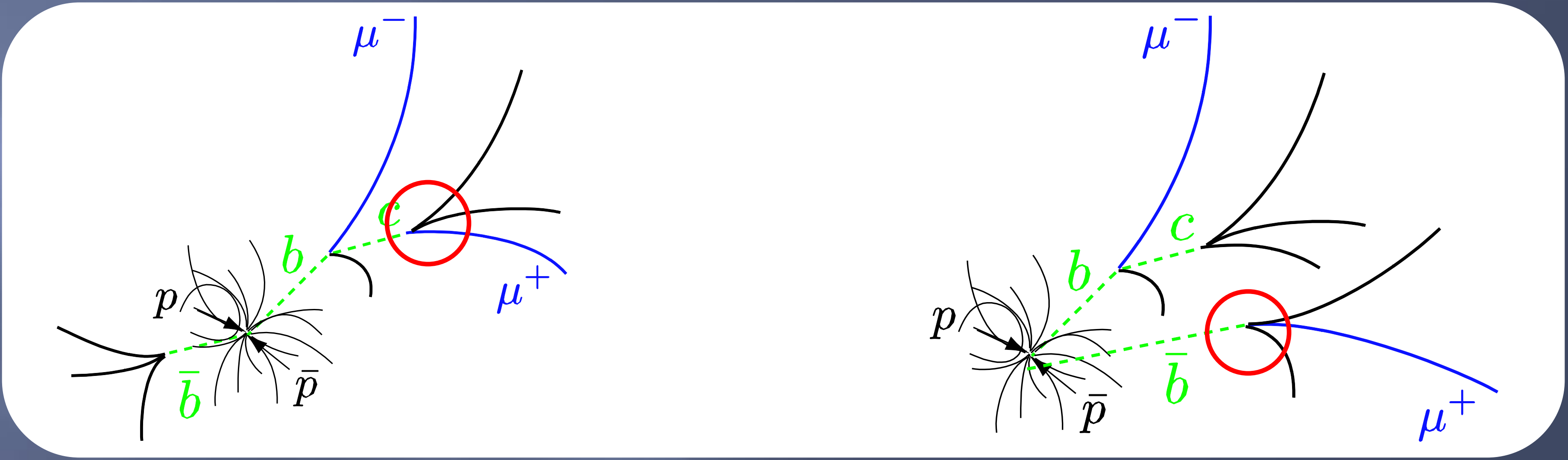
Additional Vertices



- Tertiary vertex

- Mass
- χ^2
- Impact parameter
- and others . . .

Additional Vertices

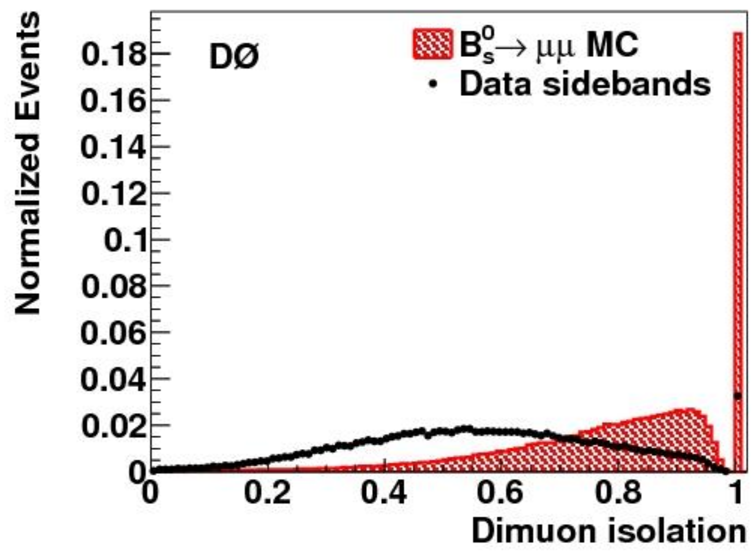


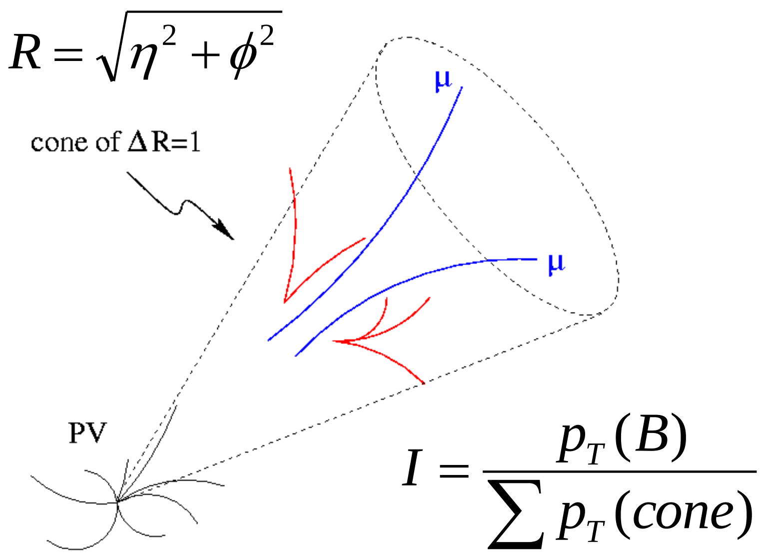
- Vertex with muons

- Mass
- χ^2
- Impact parameter
- and others . . .

Isolation

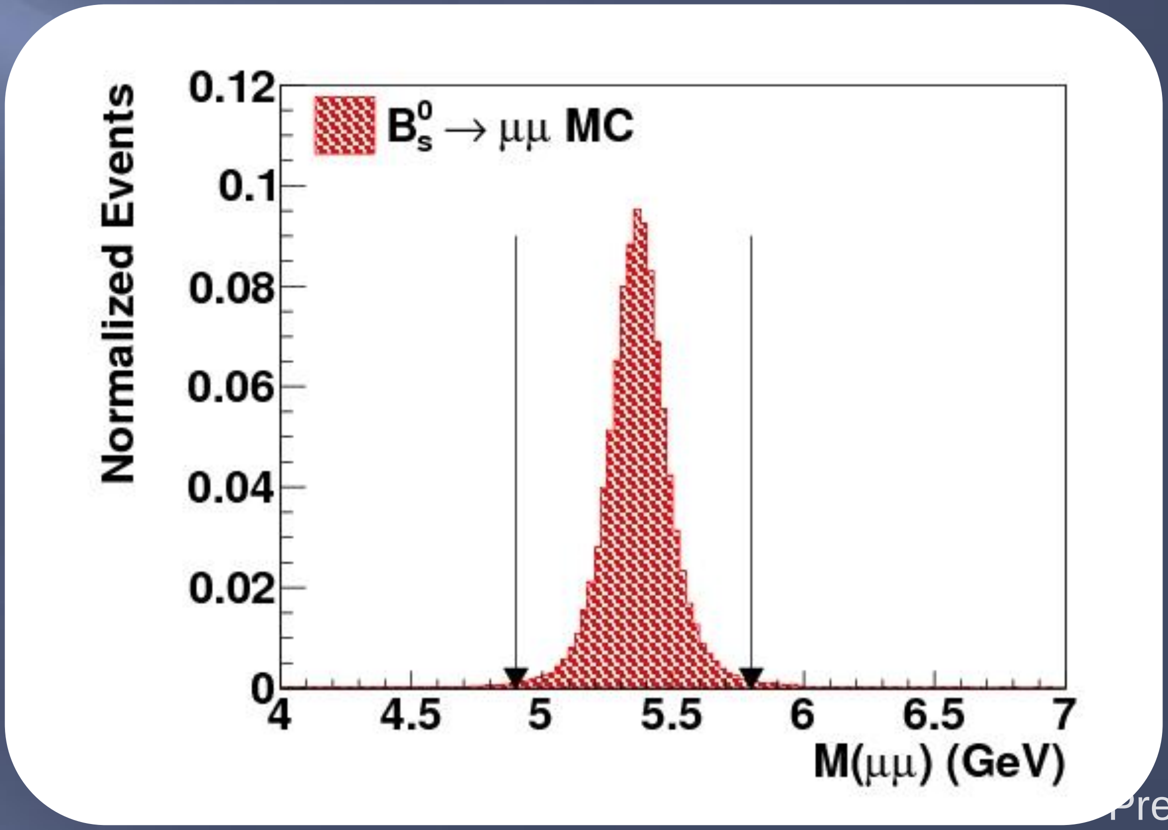
- Tracker based isolation
- Dimuon
- Individual muons





Analysis Outline

- Blinded analysis
  - Dimuon mass range of 4.0 – 7.0 GeV
  - Blinded from 4.9 – 5.8 GeV



Mass resolution
σ = 125 MeV

Analysis Outline

- Blinded analysis
  - Dimuon mass range of 4.0 – 7.0 GeV
  - Blinded from 4.9 – 5.8 GeV
- Normalization mode
  - $B^{\pm} \rightarrow J/\Psi K^{\pm}$ with $J/\Psi \rightarrow \mu^{+}\mu^{-}$

Analysis Outline

- Blinded analysis
  - Dimuon mass range of 4.0 – 7.0 GeV
  - Blinded from 4.9 – 5.8 GeV
- Normalization mode
  - $B^{\pm} \rightarrow J/\Psi K^{\pm}$ with $J/\Psi \rightarrow \mu^{+}\mu^{-}$
- New variables to fight the backgrounds

Analysis Outline

- Blinded analysis
  - Dimuon mass range of 4.0 – 7.0 GeV
  - Blinded from 4.9 – 5.8 GeV
- Normalization mode
  - $B^{\pm} \rightarrow J/\Psi K^{\pm}$ with $J/\Psi \rightarrow \mu^{+}\mu^{-}$
- New variables to fight the backgrounds
- BDT
  - Use data sidebands as background for training

Analysis Outline

- Blinded analysis
  - Dimuon mass range of 4.0 – 7.0 GeV
  - Blinded from 4.9 – 5.8 GeV
- Normalization mode
  - $B^{\pm} \rightarrow J/\Psi K^{\pm}$ with $J/\Psi \rightarrow \mu^{+}\mu^{-}$
- New variables to fight the backgrounds
- BDT
  - Use data sidebands as background for training
- Optimize cuts

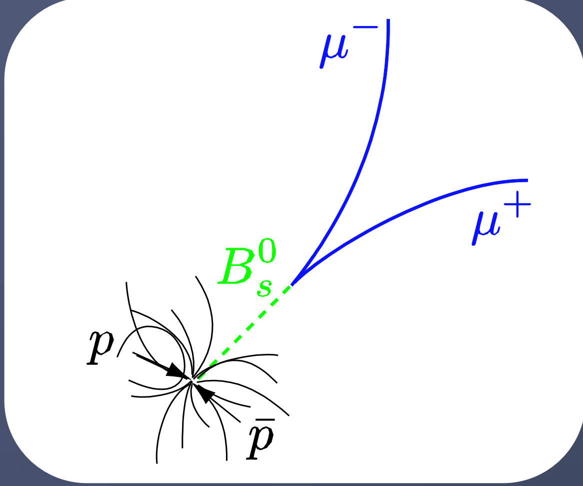
Analysis Outline

- Blinded analysis
  - Dimuon mass range of 4.0 – 7.0 GeV
  - Blinded from 4.9 – 5.8 GeV
- Normalization mode
  - $B^{\pm} \rightarrow J/\Psi K^{\pm}$ with $J/\Psi \rightarrow \mu^{+}\mu^{-}$
- New variables to fight the backgrounds
- BDT
  - Use data sidebands as background for training
- Optimize cuts
- Estimate signal and background, then limits

Analysis Outline

- Blinded analysis
  - Dimuon mass range of 4.0 – 7.0 GeV
  - Blinded from 4.9 – 5.8 GeV
- Normalization mode
  - $B^{\pm} \rightarrow J/\Psi K^{\pm}$ with $J/\Psi \rightarrow \mu^{+}\mu^{-}$
- New variables to fight the backgrounds
- BDT
  - Use data sidebands as background for training
- Optimize cuts
- Estimate signal and background, then limits
- Results

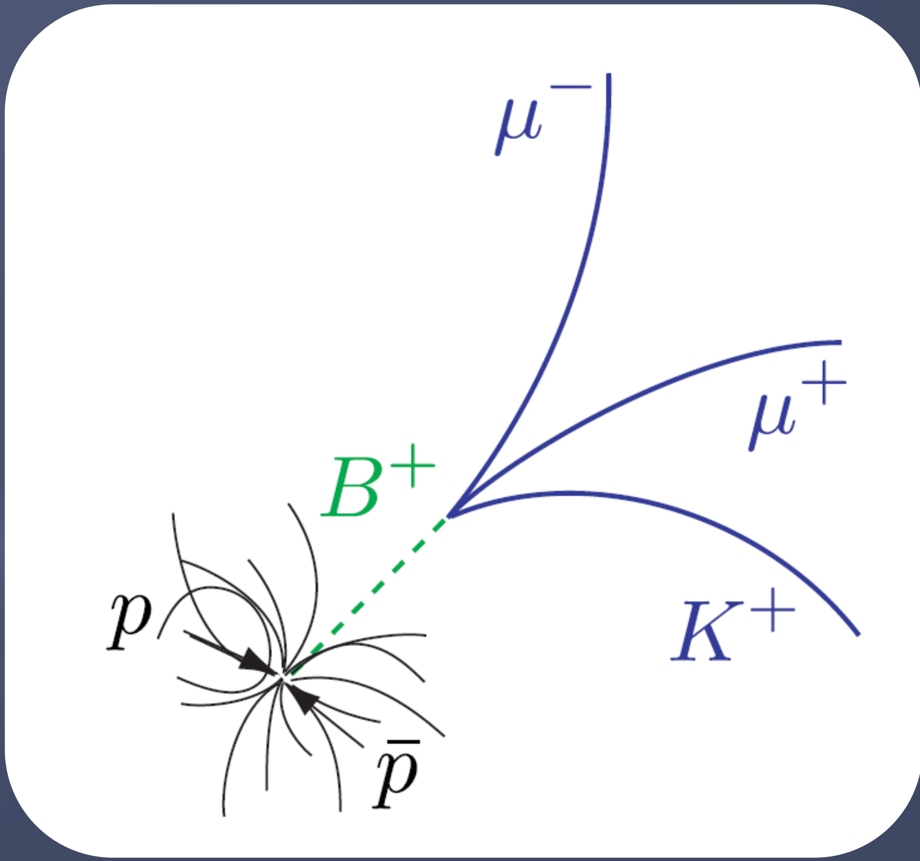
Event Selection



- Inclusive trigger strategy
- 2 oppositely charged muons
- Muon must have at least 2 hits in SMT and CFT and a match in the muon system
- Muon p_T >1.5 GeV and $|\eta|<2$
- B p_T < 20 GeV
- B vertex χ^2/dof<14
- B decay length significance >3
- Additionally in the normalization mode
  - $K^{\pm}$ must have p_T >1.0 GeV and $|\eta|<2$

Normalization Mode

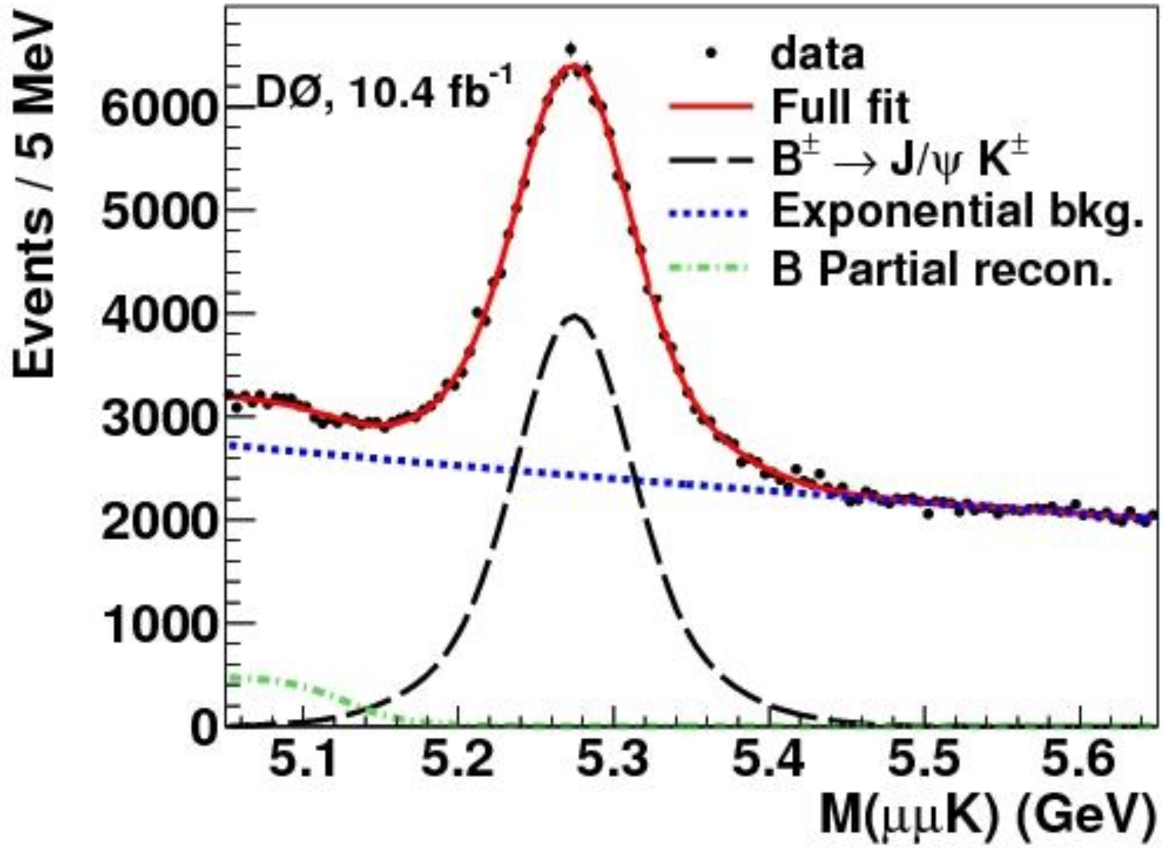
- $B^{\pm} \rightarrow J/\psi K^{\pm}$, $J/\psi \rightarrow \mu^{+}\mu^{-}$
- Find the dimuon vertex and add an additional track
- Allows cancellation of many of the systematics
- Used to validate Monte Carlo



Signal Monte Carlo

- Used PYTHIA integrated with EVTGEN for event generation
- Detector response simulated using GEANT
- Overlay data to model multiple events
- Primary $b\bar{b}$ production and gluon splitting
- Use $B^{\pm}$ to validate the B_s system
- Reweight
  - p_T
  - Isolation
  - Width of the dimuon mass
- All variables validated after reweighting

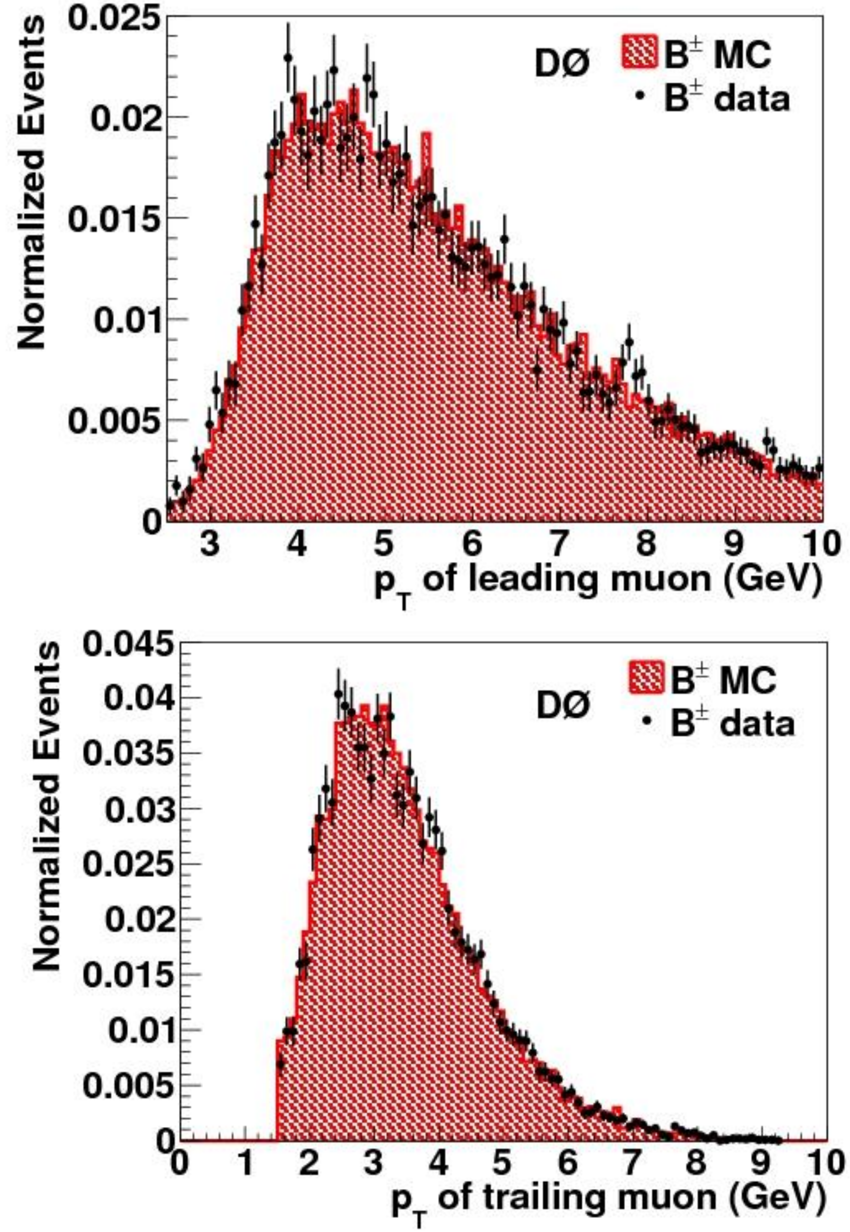
$B^{\pm}$ Data



DØ, 10.4 fb^{-1}
data
Full fit
$B^{\pm} \rightarrow J/\psi K^{\pm}$
Exponential bkg.
B Partial recon.
Events / 5 MeV
0
1000
2000
3000
4000
5000
6000
5.1
5.2
5.3
5.4
5.5
5.6
M(μμK) (GeV)

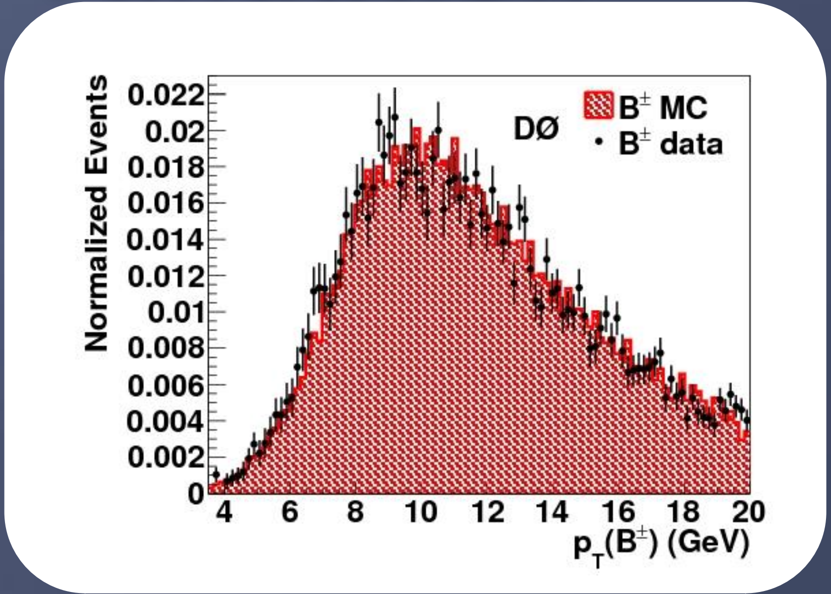
p_T Reweighting

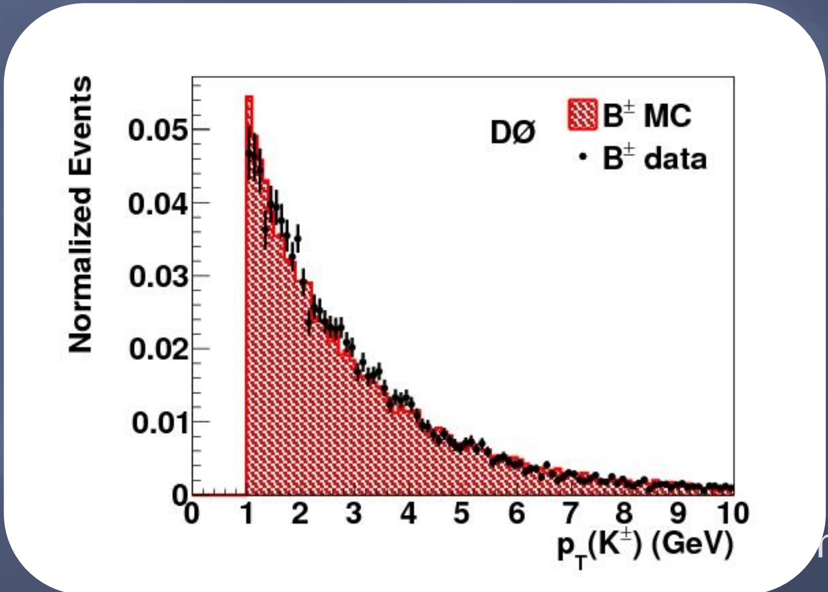
- Reproduce trigger efficiencies
  - Leading and trailing muon
- Slight reweighting above the trigger turn on for the B p_T
- Weights done for each data epoch

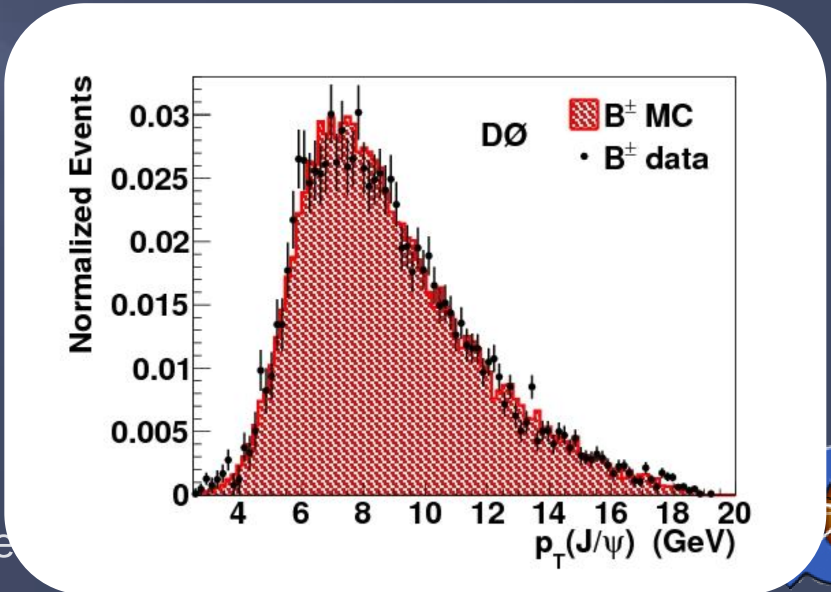


p_T Reweighting Validation

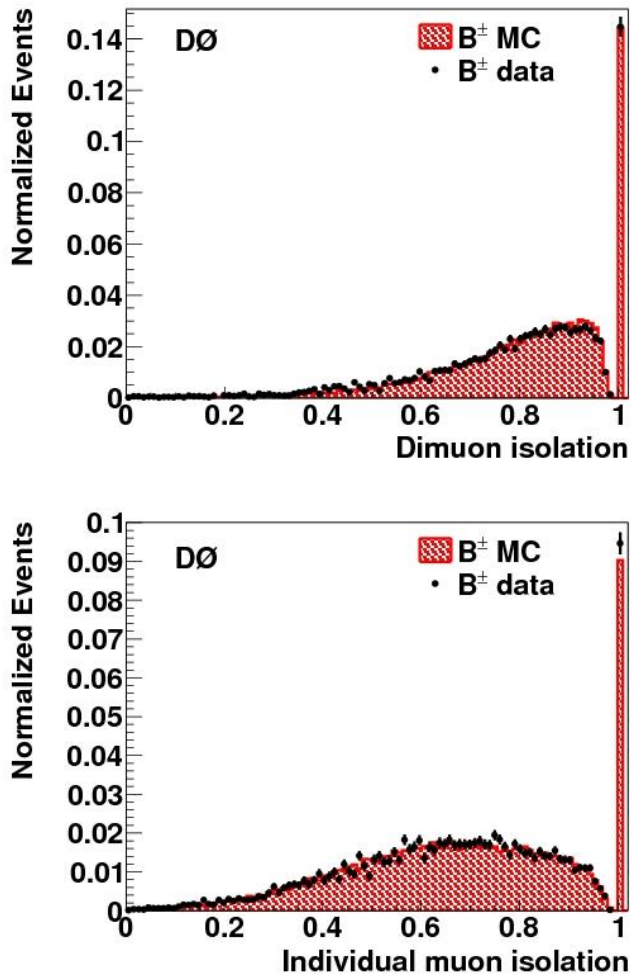
- After reweighting the muons, the other p_T distributions came out correct





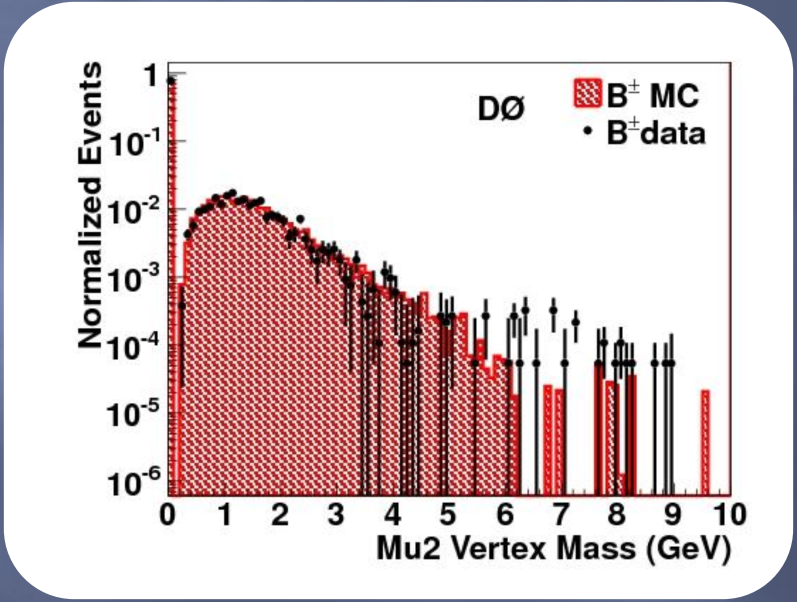


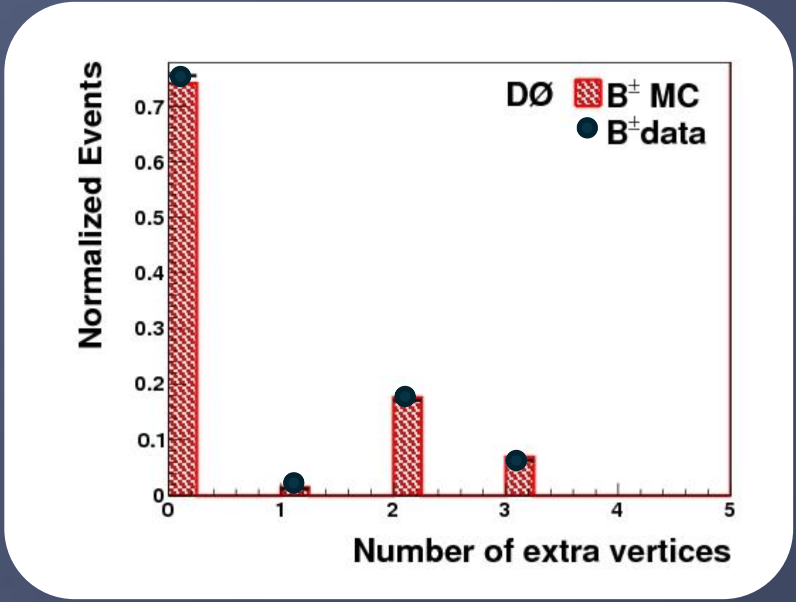
Isolation Reweighting

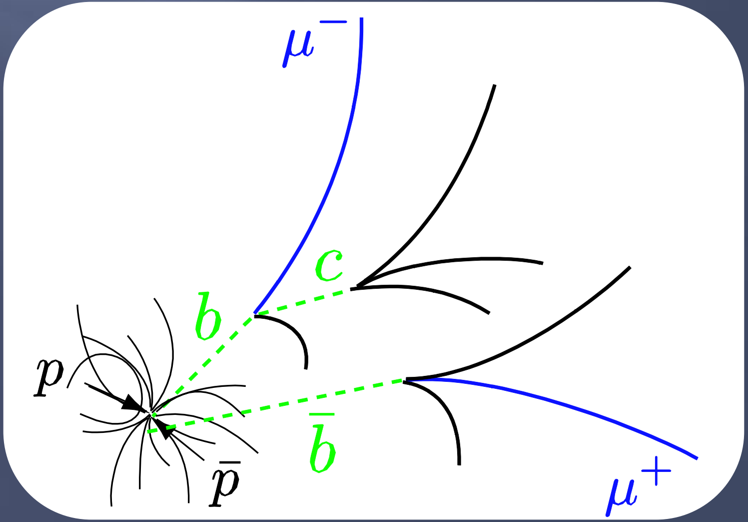


- Isolation shape matched between MC and data
- MC overestimated the number of totally isolated events (peak at 1)
- For the B, the number of events in the I=1 bin was scaled down in the MC to match the data
- After B reweighting, obtain good agreement in the individual muon isolation variable for both muons

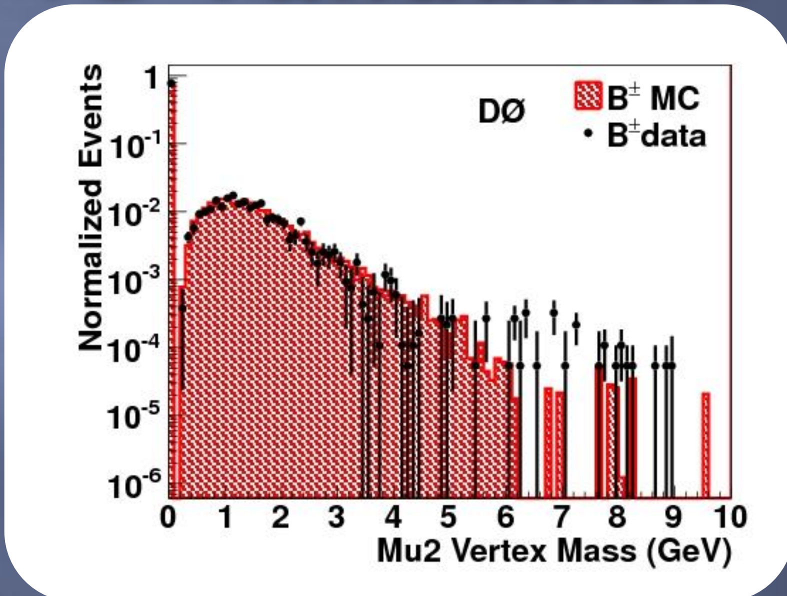
MC Validation

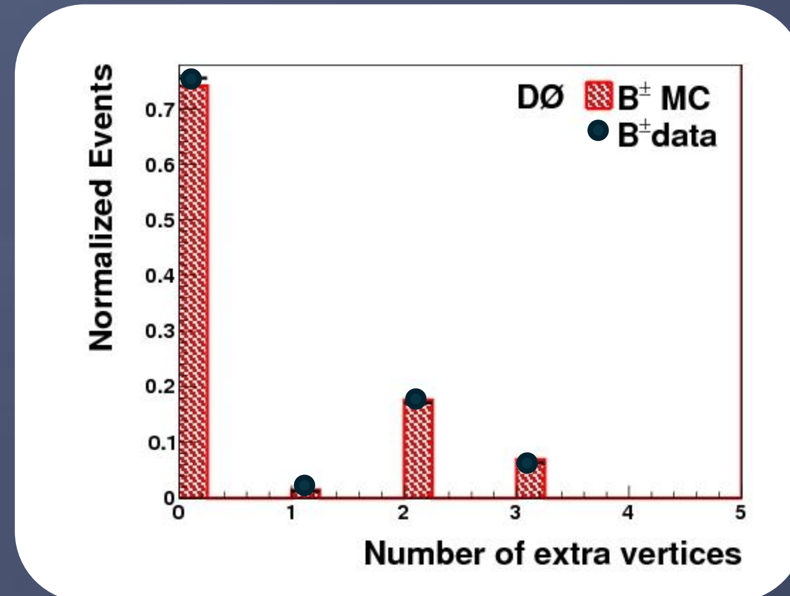


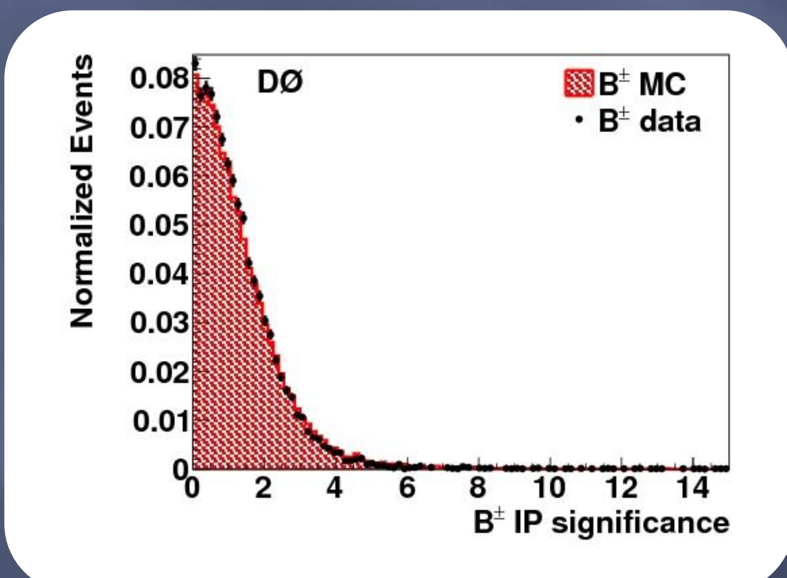


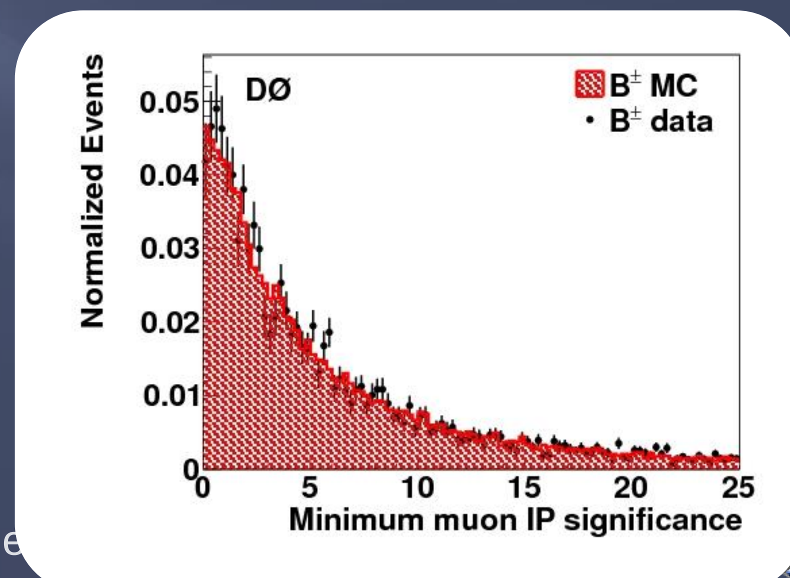


MC Validation









$$\frac{IP}{\sigma_{IP}}$$

Single Event Sensitivity

$$SES = \frac{1}{N(B^{\pm})} \times \frac{\varepsilon(B^{\pm})}{\varepsilon(B_s^0)} \times \frac{f(b \to B^{\pm})}{f(b \to B_s^0)} \times BR(B^{\pm} \to J/\psi K^{\pm}, J/\psi \to \mu^+ \mu^-)$$

- The single event sensitivity (SES) is the branching ratio at which you expect 1 event in your data sample
- The SES gives us the number of expected B_s events in the data

Single Event Sensitivity

$$SES = \frac{1}{N(B^{\pm})} \times \frac{\varepsilon(B^{\pm})}{\varepsilon(B_s^0)} \times \frac{f(b \to B^{\pm})}{f(b \to B_s^0)} \times BR(B^{\pm} \to J/\psi K^{\pm}, J/\psi \to \mu^+\mu^-)$$

- Number of $B^{\pm}$ in the data

Single Event Sensitivity

$$SES = \frac{1}{N(B^{\pm})} \times \frac{\varepsilon(B^{\pm})}{\varepsilon(B_s^0)} \times \frac{f(b \to B^{\pm})}{f(b \to B_s^0)} \times BR(B^{\pm} \to J/\psi K^{\pm}, J/\psi \to \mu^+\mu^-)$$

- Number of $B^{\pm}$ in the data
- Ratio of reconstruction efficiency for finding a $B^{\pm}$ over the efficiency for finding a B_s
  - Determined from MC with checks from data

Single Event Sensitivity

$$SES = \frac{1}{N(B^{\pm})} \times \frac{\varepsilon(B^{\pm})}{\varepsilon(B_s^0)} \times \frac{f(b \to B^{\pm})}{f(b \to B_s^0)} \times BR(B^{\pm} \to J/\psi K^{\pm}, J/\psi \to \mu^+ \mu^-)$$

- Number of $B^{\pm}$ in the data
- Ratio of reconstruction efficiency for finding a $B^{\pm}$ over the efficiency for finding a B_s
- Ratio of the fragmentation fraction of a b going to a $B^{\pm}$ versus a B_s

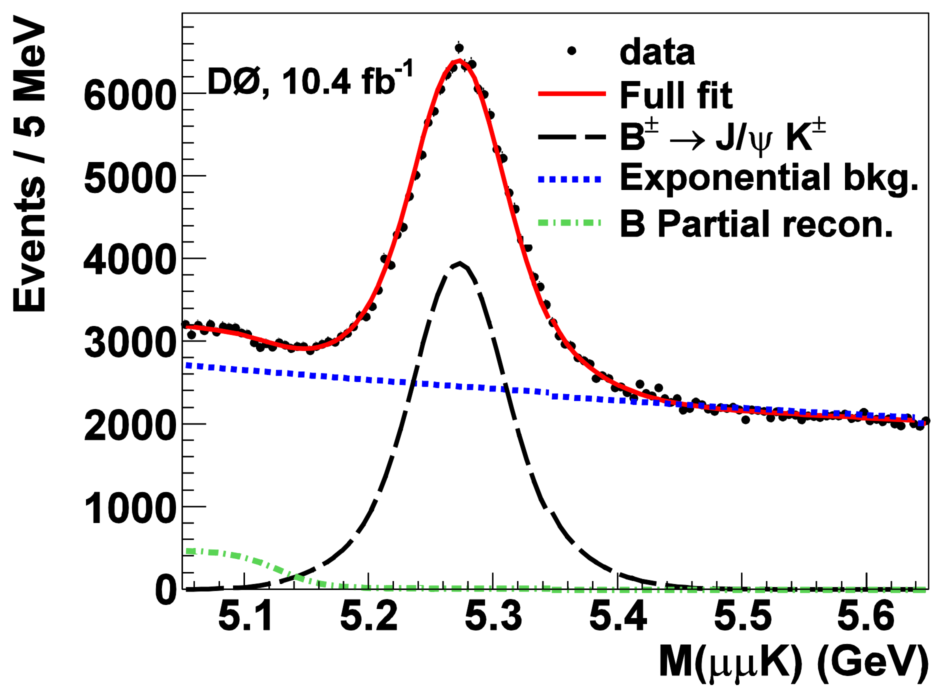
Single Event Sensitivity

$$SES = \frac{1}{N(B^{\pm})} \times \frac{\varepsilon(B^{\pm})}{\varepsilon(B_s^0)} \times \frac{f(b \to B^{\pm})}{f(b \to B_s^0)} \times BR(B^{\pm} \to J/\psi K^{\pm}, J/\psi \to \mu^+\mu^-)$$

- Number of $B^{\pm}$ in the data
- Ratio of reconstruction efficiency for finding a $B^{\pm}$ over the efficiency for finding a B_s
- Ratio of the fragmentation fraction of a b going to a $B^{\pm}$ versus a B_s
- Branching ratio of a $B^{\pm} \to J/\psi K^{\pm}$ with $J/\psi \to \mu^+\mu^-$

Determining Number of $B^{\pm}$

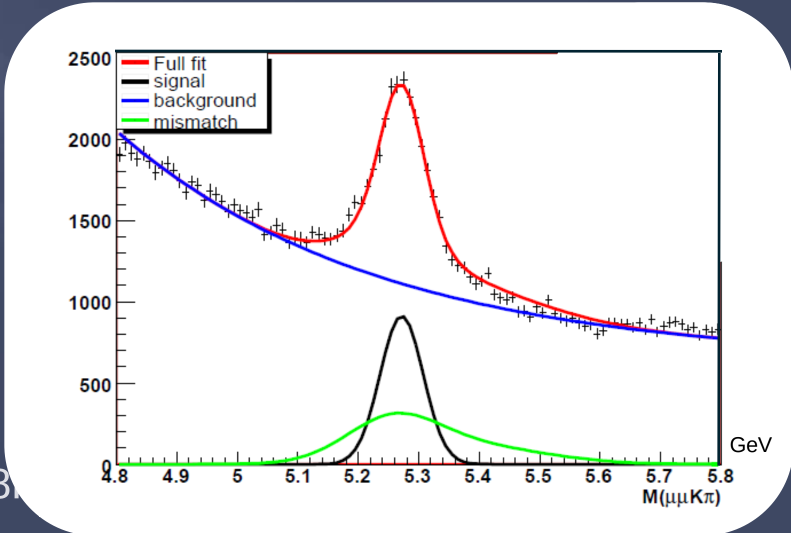
- Background fit with an exponential plus a hyperbolic tan (to model partially reconstructed B's at lower mass)
- $B^{\pm}$ fit with double Gaussian
- Systematics
  - Event selection
  - Fit function
  - p_T reweight



3^{rd} Track Efficiency Calculation

$$\frac{\varepsilon(B_d^0 \to J/\psi K^{0*}, K^{0*} \to K\pi)}{\varepsilon(B^\pm \to J/\psi K^\pm, J/\psi \to \mu^+\mu^-)} \Rightarrow \frac{\varepsilon(B^\pm \to J/\psi K^\pm, J/\psi \to \mu^+\mu^-)}{\varepsilon(B_s^0 \to \mu^+\mu^-)}$$

- Many systematics in the efficiency ratio cancel out, but the efficiency of finding the 3^{rd} track in the $B^\pm$ decay does not
- Use a 4 track system to determine this value
- Data/MC efficiency ratio for finding the 3^{rd} track is on average 0.88 ± 0.06, varies with data epoch



SES = (3.36 ± 0.29) x10^{-10}

$$SES = \frac{1}{N(B^{\pm})} \times \frac{\varepsilon(B^{\pm})}{\varepsilon(B_s^0)} \times \frac{f(b \to B^{\pm})}{f(b \to B_s^0)} \times BR(B^{\pm} \to J/\psi K^{\pm}, J/\psi \to \mu^+ \mu^-)$$

- Number of $B^{\pm}$
  - (87.4 ± 3.0) x10^3
- Efficiency of finding $B^{\pm}$ and B_s
  - (13.0 ± 0.5)%
- Fragmentation ratio
  - HFAG 2012
  - 1/(0.263 ± 0.017)
- Branching ratio
  - PDG
  - (6.01 ± 0.21) x10^{-5}

The efficiencies were determined for each data epoch individually, so the value given is an average.

SES = (3.36 ± 0.29) x10^{-10}

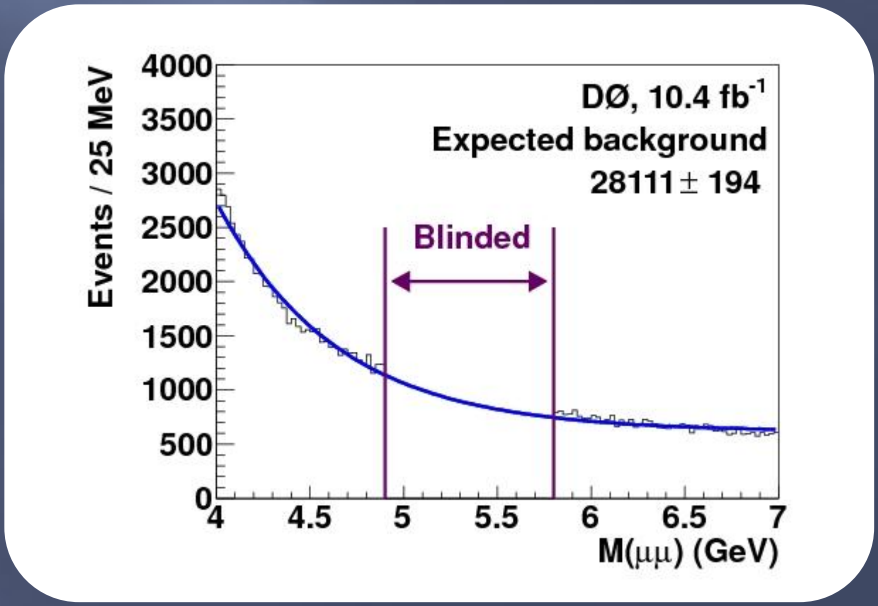
$$SES = \frac{1}{N(B^{\pm})} \times \frac{\varepsilon(B^{\pm})}{\varepsilon(B_s^0)} \times \frac{f(b \to B^{\pm})}{f(b \to B_s^0)} \times BR(B^{\pm} \to J/\psi K^{\pm}, J/\psi \to \mu^+\mu^-)$$

- Number of $B^{\pm}$
  - (87.4 ± 3.0) x10^3
- Efficiency of finding $B^{\pm}$ and B_s
  - (13.0 ± 0.5) %
- Fragmentation ratio
  - HFAG 2012
  - 1/(0.263 ± 0.017)
- Branching ratio
  - PDG
  - (6.01 ± 0.21) x10^{-5}

Expect 10.4 ± 1.1 $B_s \to \mu^+\mu^-$ events in the blinded region of the data (before BDT cuts)

Recap

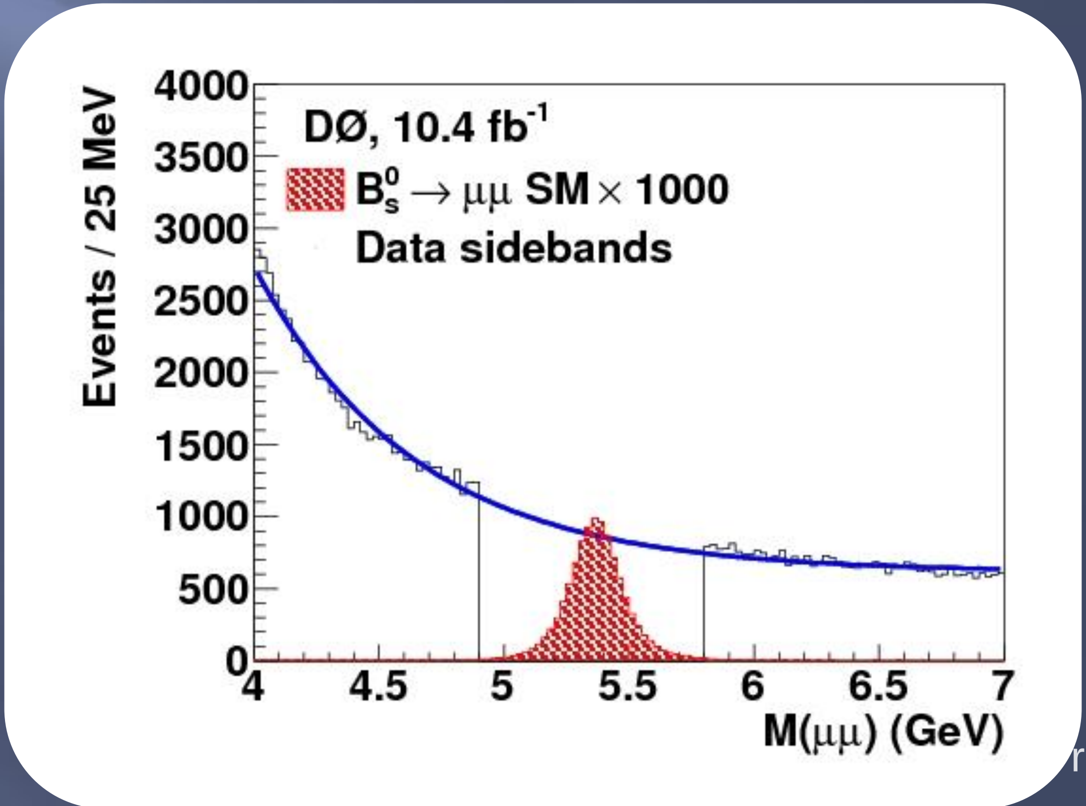
- Selected our data
- Validated our MC
- Determined the expected number of B_s
- But we are trying to find ~10 events in this



Dimuon mass distribution from data fit with an exponential plus a constant to determine the number of events in the blinded region.

Boosted Decision Tree

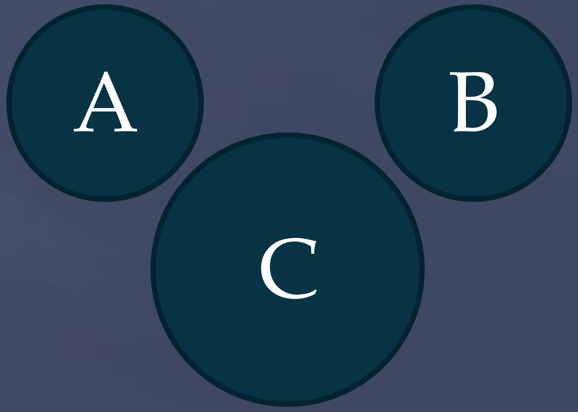
- Use a boosted decision tree (BDT) from TMVA
  - Data sidebands are used as background
  - B_s MC is used as signal



Mass resolution
σ = 125 MeV

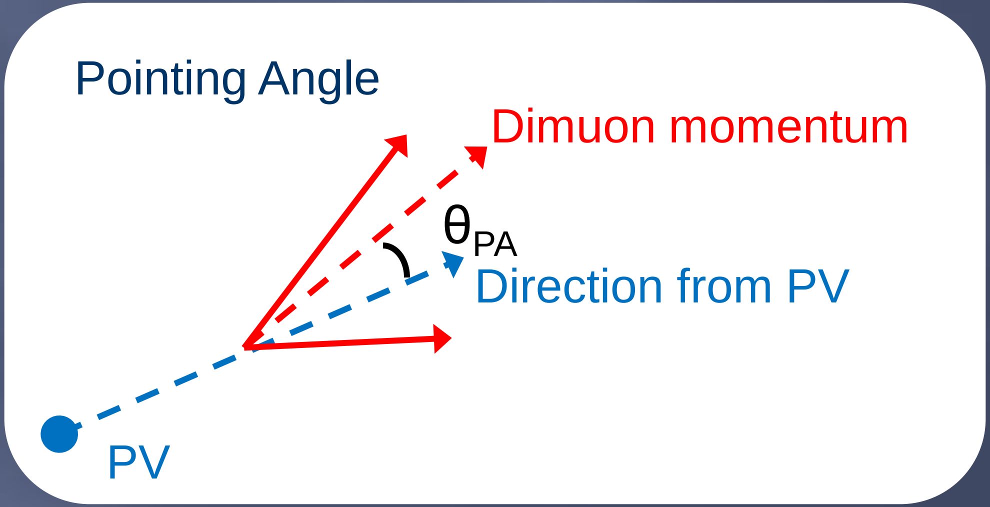
Boosted Decision Tree

- Use a boosted decision tree (BDT) from TMVA
  - Data sidebands are used as background
  - B_s MC is used as signal
- The data sidebands and the MC are separated into 3 samples
  - A (25%) for training the BDT
  - B (25%) for optimizing the cuts
  - C (50%) for estimating the results



Optimizing the BDT(s)

- Make additional requirements before BDT training
  - Cosine of 2D pointing angle > 0.95
  - Cosine of 3D pointing angle > 0.90
  - Dimuon p_T > 5 GeV



Cuts reduce background by 96% and keep 78% signal.

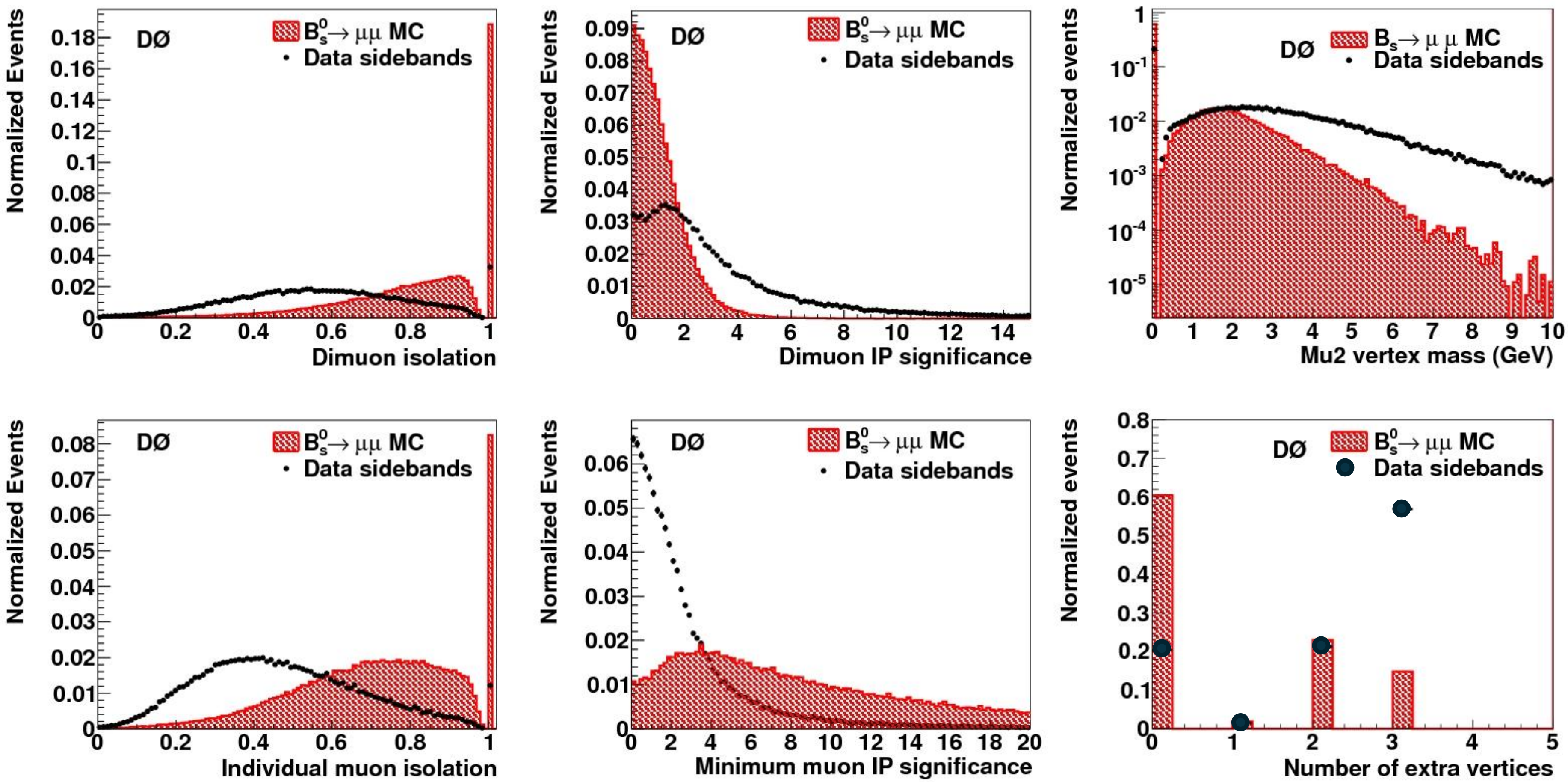
Optimizing the BDT(s)

- Make additional requirements before BDT training
  - Cosine of 2D pointing angle > 0.95
  - Cosine of 3D pointing angle > 0.90
  - Dimuon p_T > 5 GeV
- Use 2 different BDTs to discriminate against the 2 types backgrounds
  - Sequential decay
  - Double b decay
- 30 variables

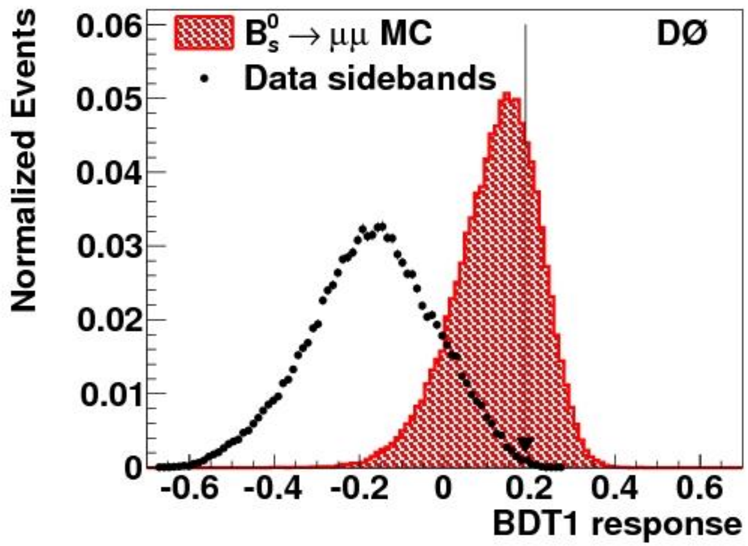
Most Important BDT Variables

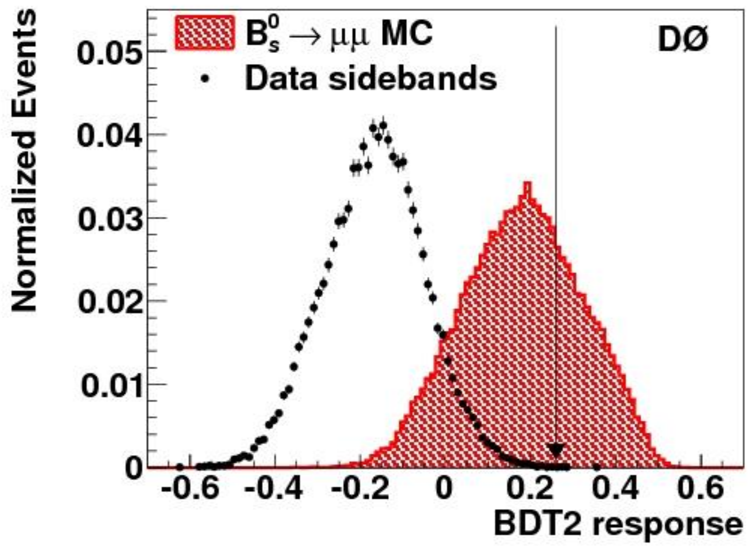
- Dimuon kinematics and topology
  - p_T, decay length, pointing angle, vertex χ^2, etc.
- Individual muon kinematics
  - p_T, impact parameter, ϕ difference, etc.
- Isolation
  - Dimuon, leading muon, and trailing muon
- Additional vertices
  - Mass, vertex χ^2, pointing angle, etc.

BDT Variables



BDT Response





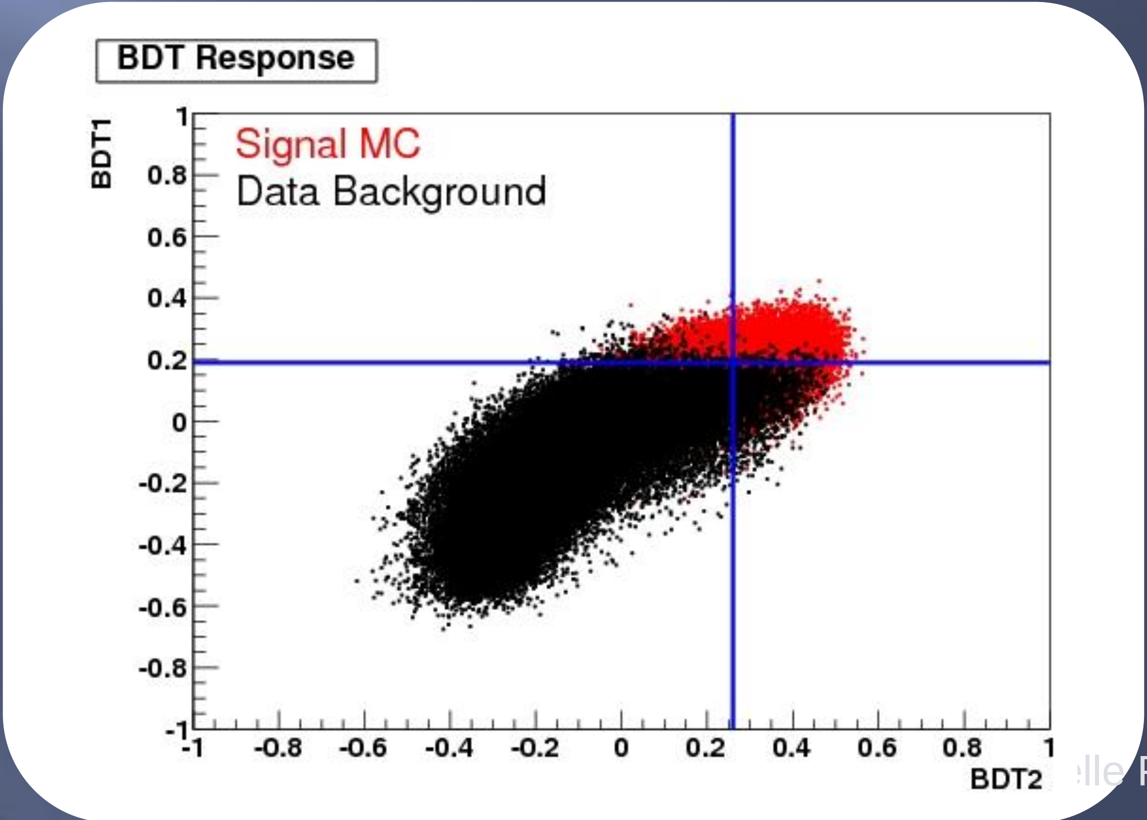
- BDT 1
- Sequential decay

- BDT 2
- Double b decay

Even though the "background" and "signal" samples are equivalent at this stage in the analysis, note that the B_s MC represents about 10 events while the data sidebands represent 10^5 events.

BDT Response Cut

- Coarse cuts optimization, maximize $S/\sqrt{(S+B)}$
- Final cuts chosen to optimize the expected limit

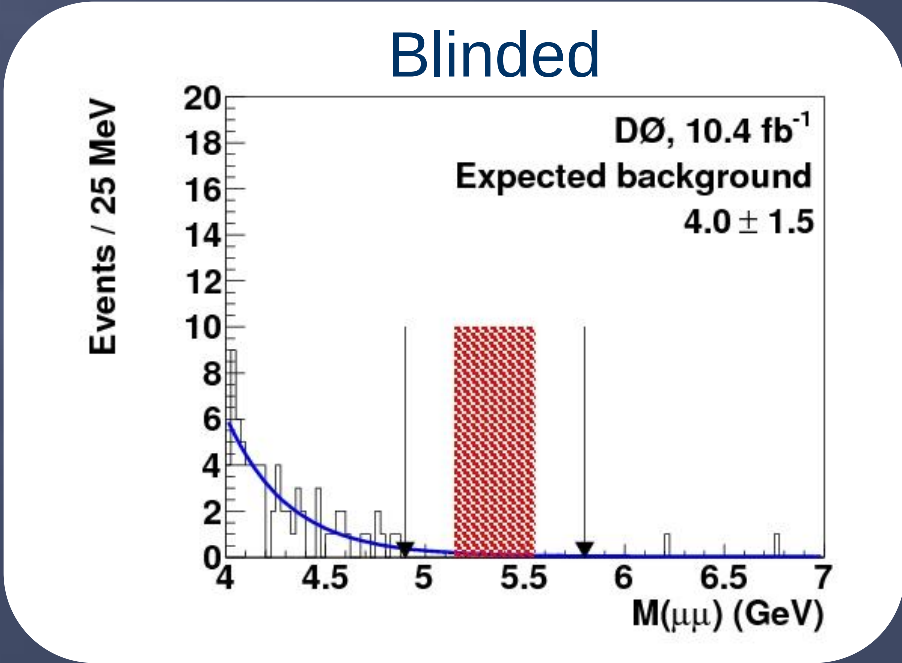


Other Backgrounds

- $B_s \rightarrow KK$ peaking background
  - K decays in flight after the tracker faking a good muon
  - $K \rightarrow \mu$ fake rate determined from $B \rightarrow \mu D^0 X$ with $D^0 \rightarrow K\pi$
  - Other $B_s \rightarrow hh$ contributions found to be negligible due to low fake rate and lower branching fractions
- $B_d \rightarrow \mu\mu$
  - Mass resolution not sensitive enough to distinguish B_d from B_s, but assume negligible B_d contribution
  - SM $BR(B_d \rightarrow \mu\mu) = (1.1 \pm 0.1) \times 10^{-10}$

Expected Signal and Background

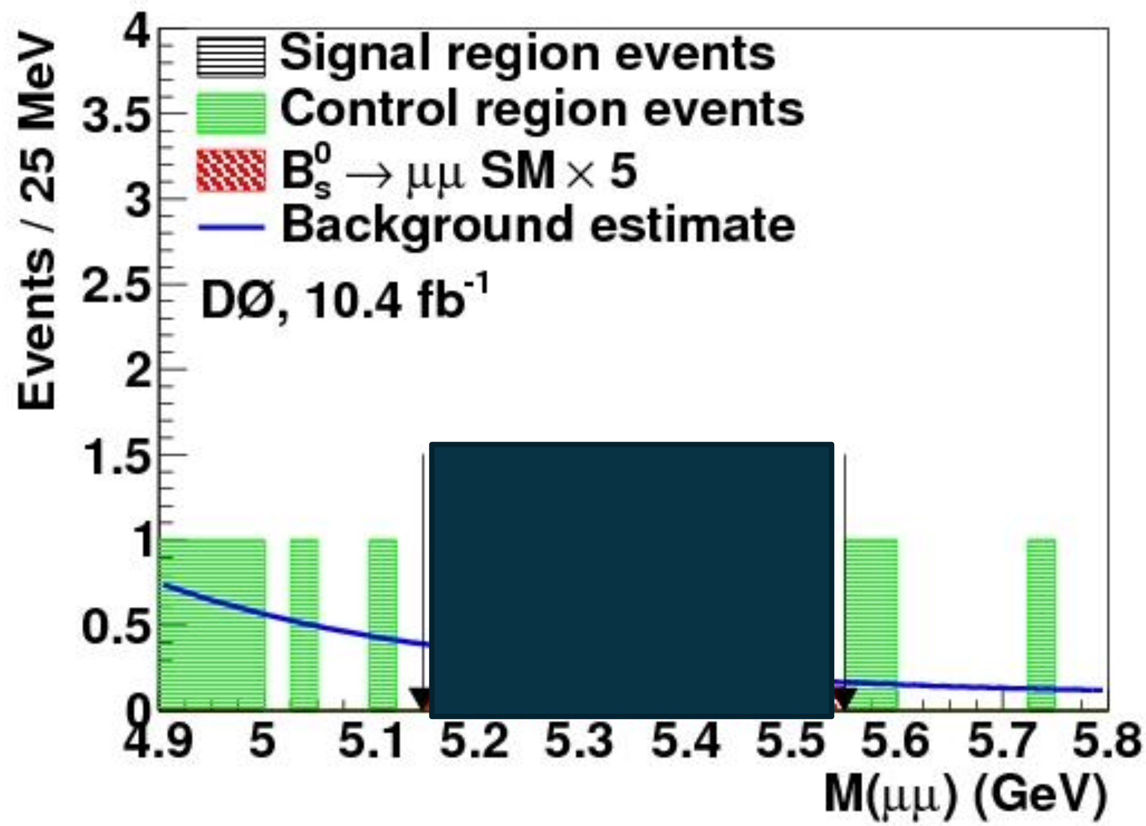
- Total Expected Background of 4.3 ± 1.6
  - 4.0 ± 1.5 ± 0.6 dimuon background events
  - 0.3 ± 0.1 $B_s \rightarrow KK$ peaking background events
- Expected SM Signal 1.23 ± 0.13
- Expected 95% C.L. upper limit on the branching ratio $BR(B_s \rightarrow \mu^+\mu^-) < 23 \times 10^{-9}$ using modified frequentist method*



*T. Junk, Nucl. Instrum. and Meth. in Phys. Res. A 434, 435 (1995).

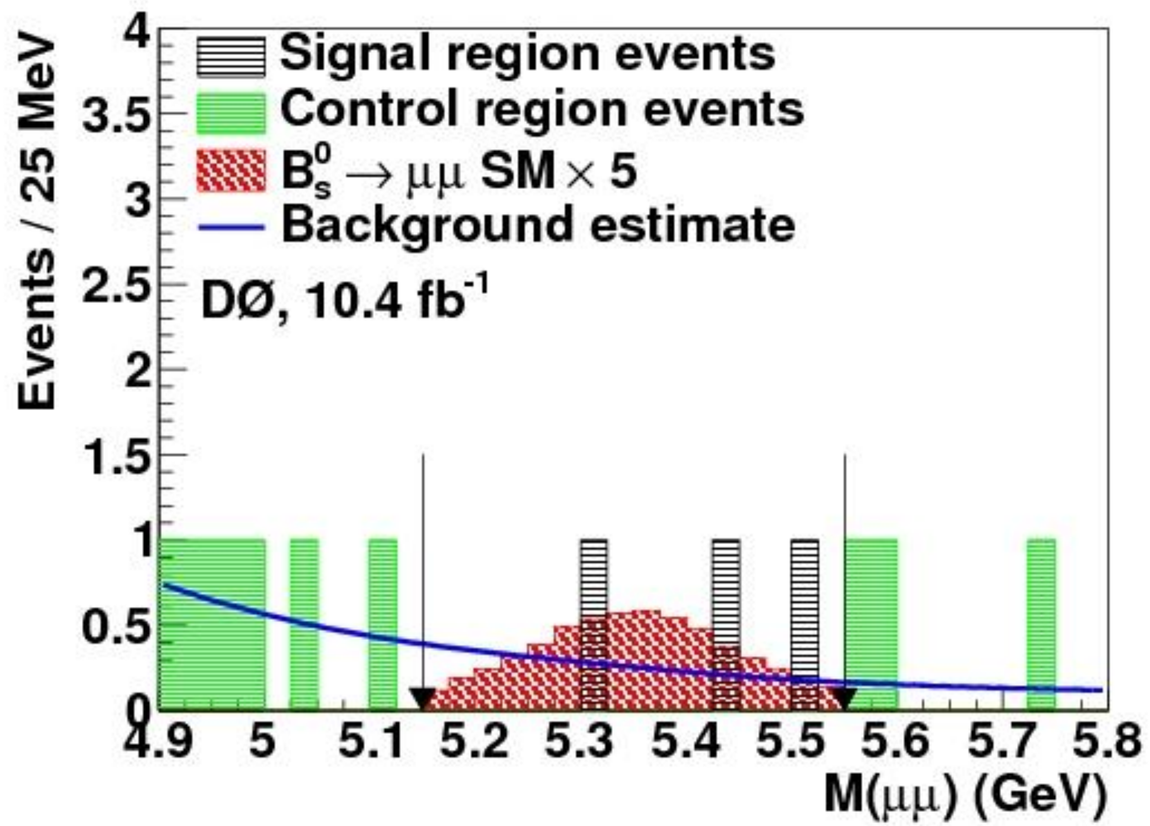
Unblinding

- As a cross check, looked at the control region first
- Expected 6.7 ± 2.6 background events
- Observe 9 events



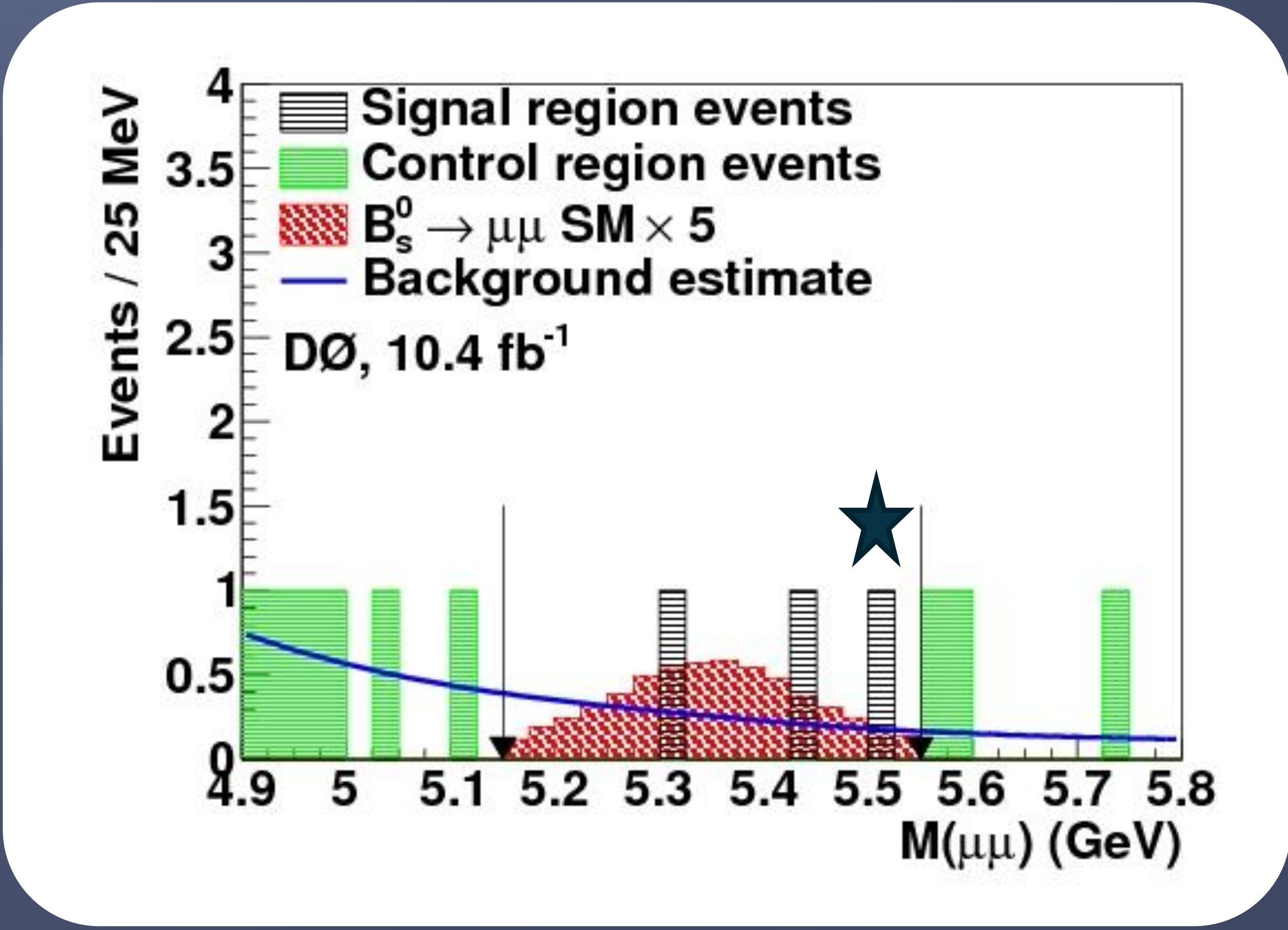
Unblinding

- Expected 4.3 ± 1.6 background events
- Expected 1.23 ± 0.13 signal events
- Observed 3 events in the signal region

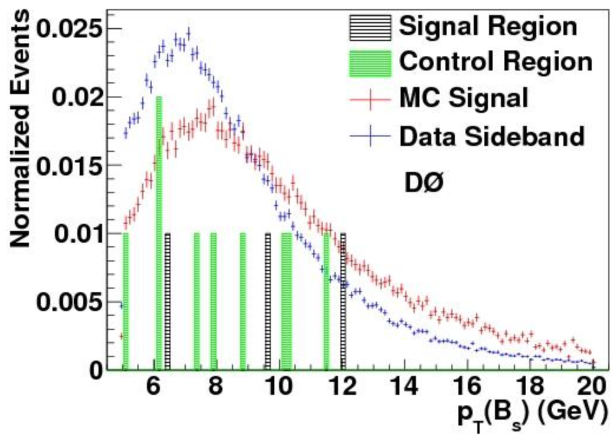


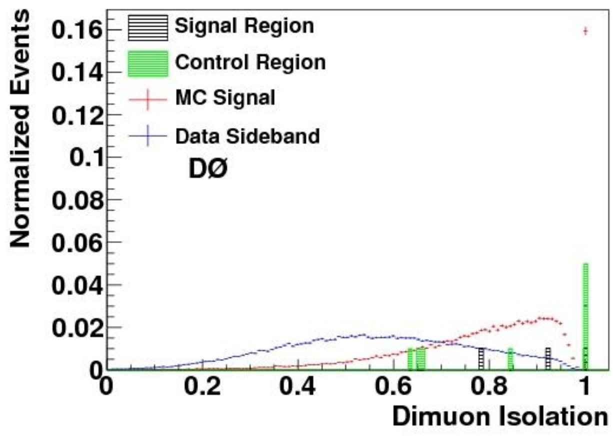
Unblinding

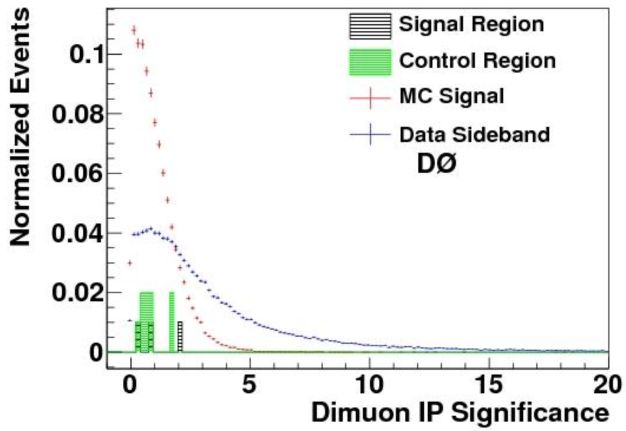
- Expected 4.3 ± 1.6 background events
- Expected 1.23 ± 0.13 signal events
- Observed 3 events in the signal region

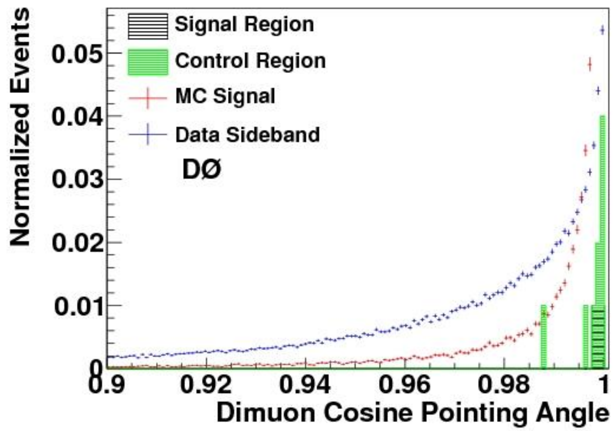


Characteristics of Candidates



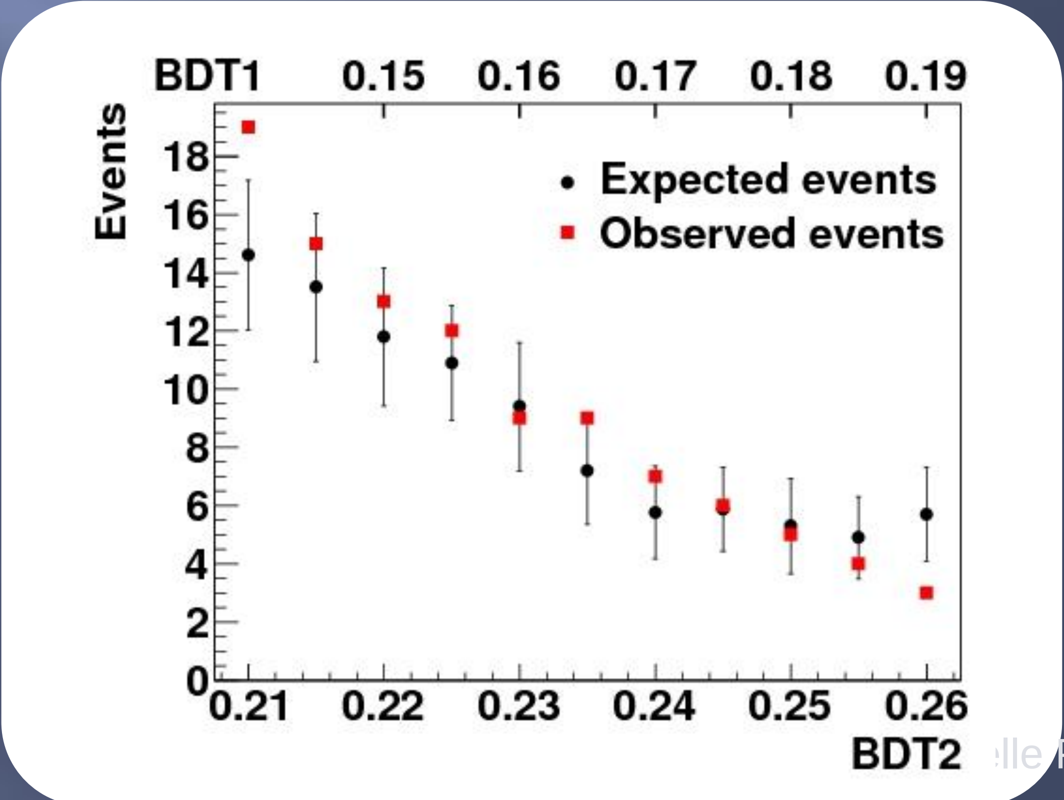






Cross Checks

- Compared the number of expected events to the number of observed events as a function of BDT response cut

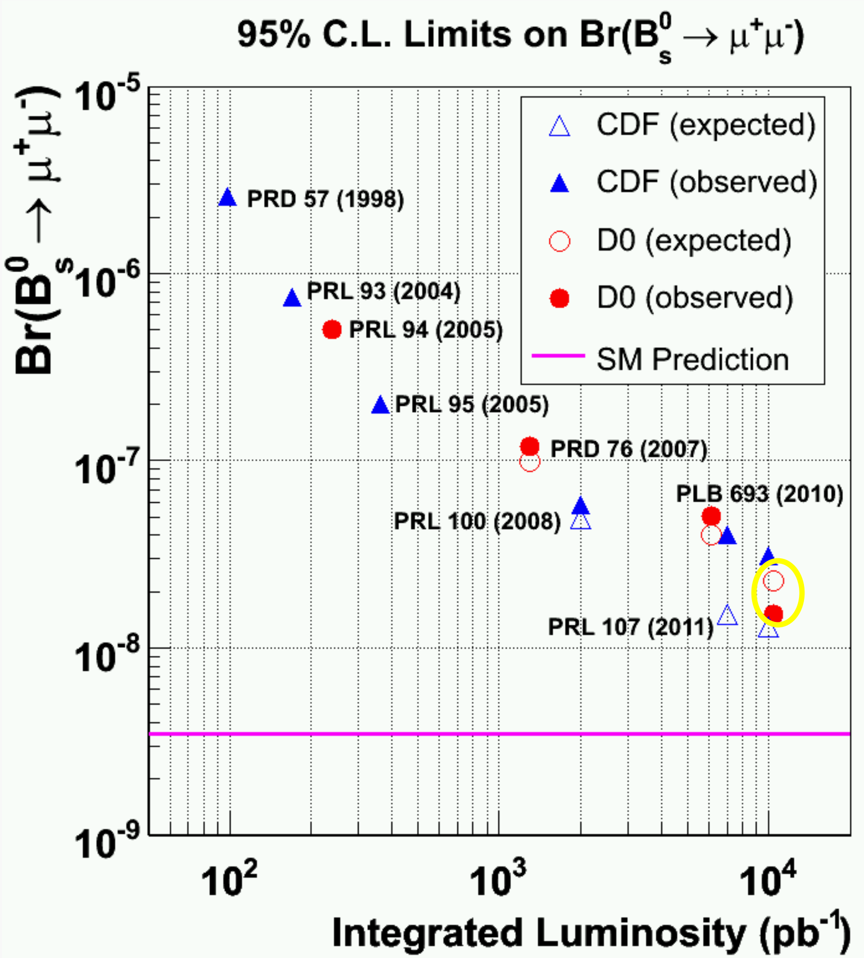


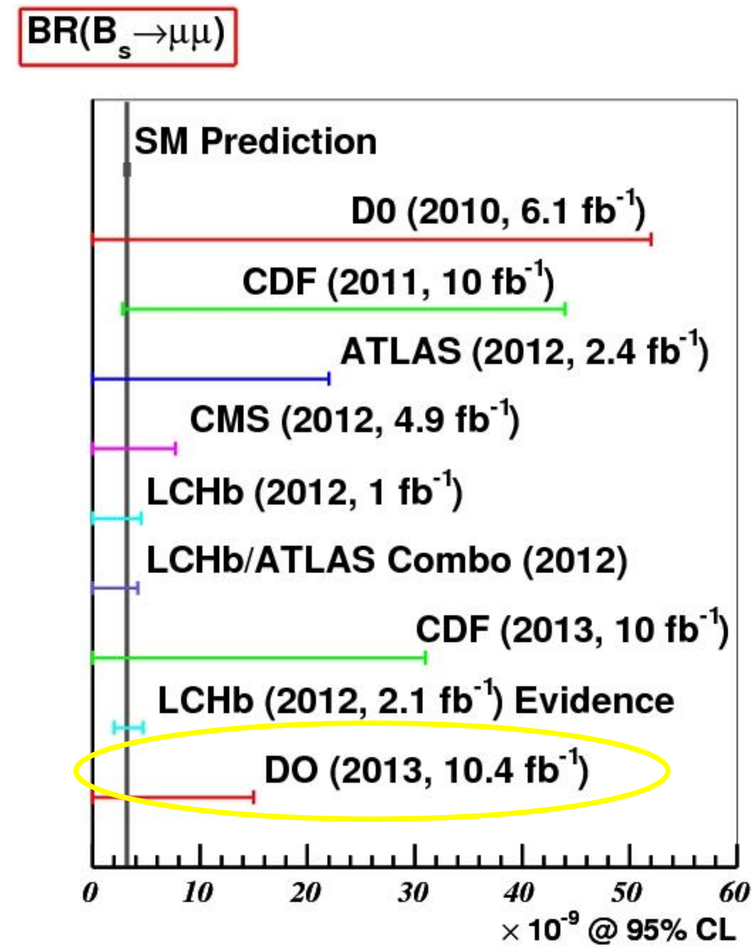
Result

- Expected
  - SM Signal: 1.23 ± 0.13
  - Background: 4.3 ± 1.6
  - Limit: $BR(B_s \to \mu^+\mu^-) = 23 \times 10^{-9}$
- Observed 3 events setting a 95% C.L. limit on the branching fraction of

$$BR(B_s^0 \to \mu^+\mu^-) < 15 \times 10^{-9}$$

World Measurements





Summary

- Searched for the decay of $B_s \to \mu^+\mu^-$
- Set a limit on the branching ratio of 15×10^{-9}
- Improved upon previous D0 results by a factor of 3.4 and a factor of 1.7 better than the $\sqrt{\mathcal{L}}$ for the expected limit
  - Created new variables to distinguish signal from background
  - Used separate BDTs for the different types of backgrounds
- Best observed Tevatron limit

Summary

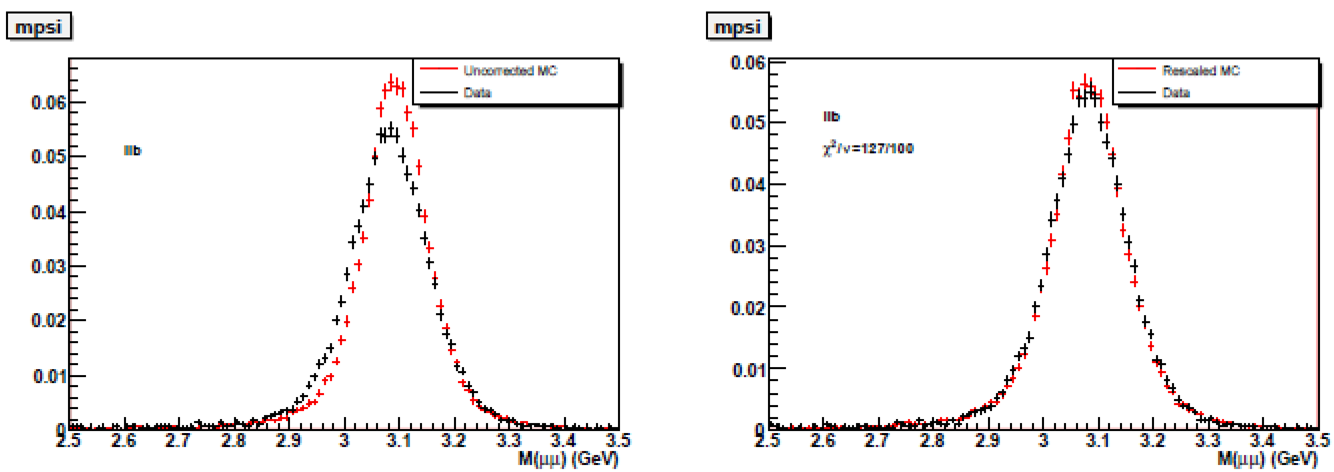
- Searched for the decay of $B_s \to \mu^+\mu^-$
- Set a limit on the branching ratio of 15×10^{-9}
- Improved upon previous D0 results by a factor of 3.4 and a factor of 1.7 better than the $\sqrt{\mathcal{L}}$
  - Created new variables to distinguish signal from background
  - Used separate BDTs for the different types of backgrounds
- Best observed Tevatron limit

Thanks!

Backup

Mass Width Reweight

- Width of the dimuon mass was reweighted in the MC to match the data using $J/\psi \rightarrow \mu^+\mu^-$
- The correction was then scaled to the B_s mass
- Approximately 17 MeV shift and 13% stretch



Optimizing the BDT

- Types of BDT
- BDT parameters
  - Ex. Number of trees
- Variables used in BDT
- Using different BDTs for the different backgrounds
- Additional cuts before training the BDT

BDT Variables

- $ptmumu$: p_T of the dimuon system.
- $cxymumu$: Cosine of the dimuon pointing angle, calculated using information only in the transverse plane.
- $c3dmumu$: Cosine of the dimuon pointing angle using 3D information.
- $lxymumu$: Dimuon decay length, calculated using information only in the transverse plane.
- $l3dmumu$: Dimuon decay length using 3D information.
- $sigxymumu$: Dimuon decay length significance.
- $ip1mumu$: Dimuon impact parameter, calculated using information only in the transverse plane.
- $ipsig1mumu$: Dimuon impact parameter significance.
- $chi2mumu$: χ^2 of the dimuon vertex.
- $ptxysqmumu$: Square of the dimuon momentum component perpendicular to the line from the primary vertex to the dimuon vertex, calculated in the transverse plane.
- $pt3dsqmumu$: Same as $ptxysqmumu$ excepted calculated using 3D information.
- iso: Standard isolation variable $I=p_T(\mu \mu)$/[$p_T(cone)$+$p_T(\mu \mu)]$ in R~=~1 cone.
- $isomu1$: Same as iso, but defined with respect to the leading muon rather than the dimuon direction.
- $isomu2$: Isolation defined with respect to the trailing muon direction.
- $isosum$: Sum of the individual muon isolation. $isomu1+isomu2$
- $ptmu1$: p_T of the leading muon.
- $ptmu2$: p_T of the trailing muon.
- $ip1mu1$: Leading muon impact parameter.
- $ip1mu2$: Trailing muon impact parameter.
- $ipsig1mu1$: Leading muon impact parameter significance.
- $ipsig1mu2$: Trailing muon impact parameter significance.
- $deltaphi$: Difference in azimuthal angles between the two muons.
- $ipsigless$: Smaller of the two impact parameters of the two muons.
- $cxyNew$: Cosine of the pointing angle. %need to look up
- $mTer$: Invariant mass of the tracks associated with an additional vertex that does not include either muon.
- $mTermu1$: Invariant mass of the tracks associated with an additional vertex that includes the leading muon.
- $mTermu2$: Invariant mass of the tracks associated with an additional vertex that includes the trailing muon.
- $chi2mu1iso$: χ^2 of the vertex of tracks with the leading muon.
- $chi2mu2iso$: χ^2 of the vertex of tracks with the trailing muon.
- $cxymu2iso$: Cosine of the pointing angle for the vertex of tracks with the trailing muon.

Signal Region

- Blinded dimuon mass range from 4.9 – 5.8 GeV is a 3σ window about the B_s mass peak
- Optimized the signal region 1.6σ window
- Control region from 4.9 – 5.15 & 5.55 – 5.8 GeV
- Signal region from 5.15 – 5.55 GeV

Expected Signal and BG

- Apply both BDT trainings to independent sample C to estimate the number of signal and background events after BDT response cuts.
- Determine estimated background with a loglikelihood fit of exponential + constant to data sidebands and extrapolate into the signal region.
- Took ratio of MC events before and after BDT cuts to determine efficiency for signal to pass BDT cuts.
- Then multiplied SES and efficiency to determine final number of expected signal events.