Measurement of $B_s^0 \to \mu^+\mu^-$ with CMS

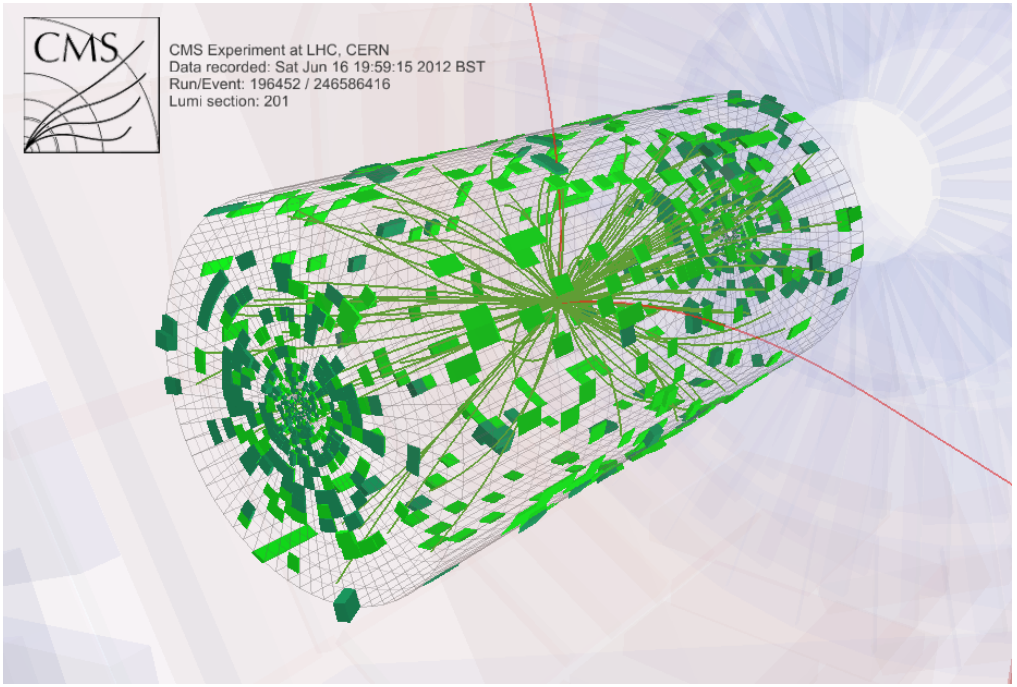
Urs Langenegger
(PSI)

for the CMS collaboration

EPS HEP Stockholm
2013/07/19

- Introduction
  - ▷ motivation and methodology
  - ▷ detector
- Analysis
  - ▷ selection
  - ▷ validation
- Results
  - ▷ $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$

arxiv:1307.5025, subm. to PRL

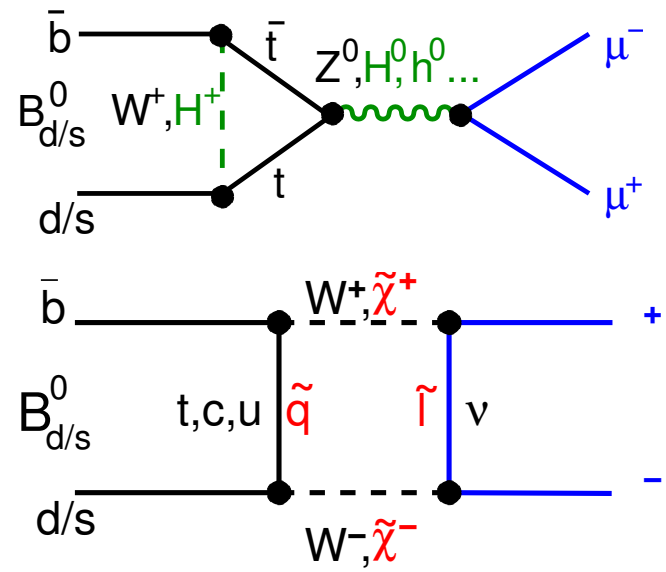


Introduction

- Decays highly suppressed in Standard Model
  - ▷ effective FCNC, helicity suppression
  - ▷ SM (decay-time integrated) expectation:

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (3.57 \pm 0.30) \times 10^{-9}$$
$$\mathcal{B}(B^0 \to \mu^+\mu^-) = (1.07 \pm 0.10) \times 10^{-10}$$

(Buras, *et al.*, 2012)



  - ▷ Cabibbo-enhancement $(|V_{ts}| > |V_{td}|)$ of $B_s^0 \to \mu^+\mu^-$ over $B^0 \to \mu^+\mu^-$ only in MFV models
  - ▷ high sensitivity to models with extended Higgs-boson sectors

- Complete re-analysis of entire 2011+2012 dataset
  - ▷ improved muon identification (BDT)
  - ▷ new and improved variables
  - ▷ MVA selection (BDT) plus unbinned maximum-likelihood fit

$\Rightarrow$ (Re-)Blind ($\approx$ half of) data for $5.2 < m_{\mu\mu} < 5.45$ GeV

  - ▷ selection development and choice of interpretation methodology

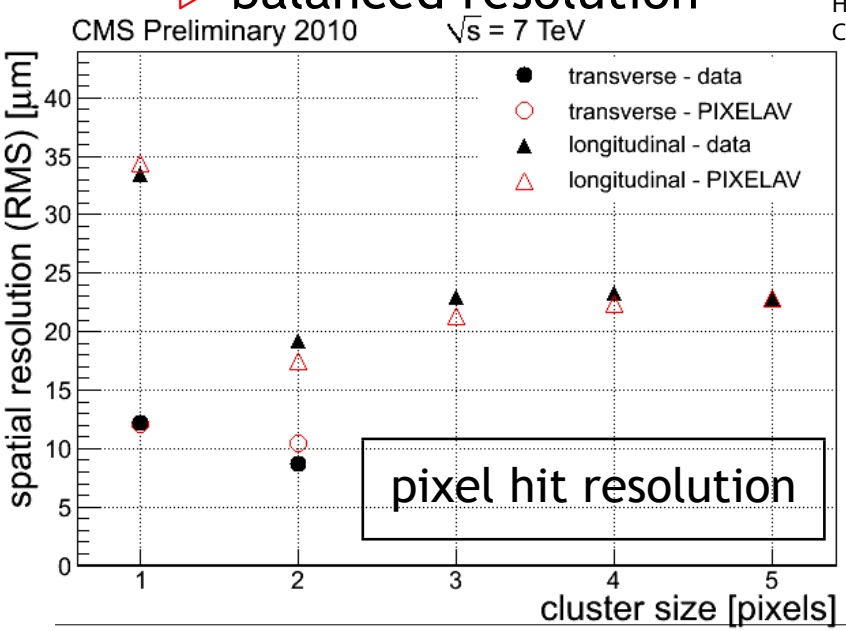
JINST 3, S08004 (2008)

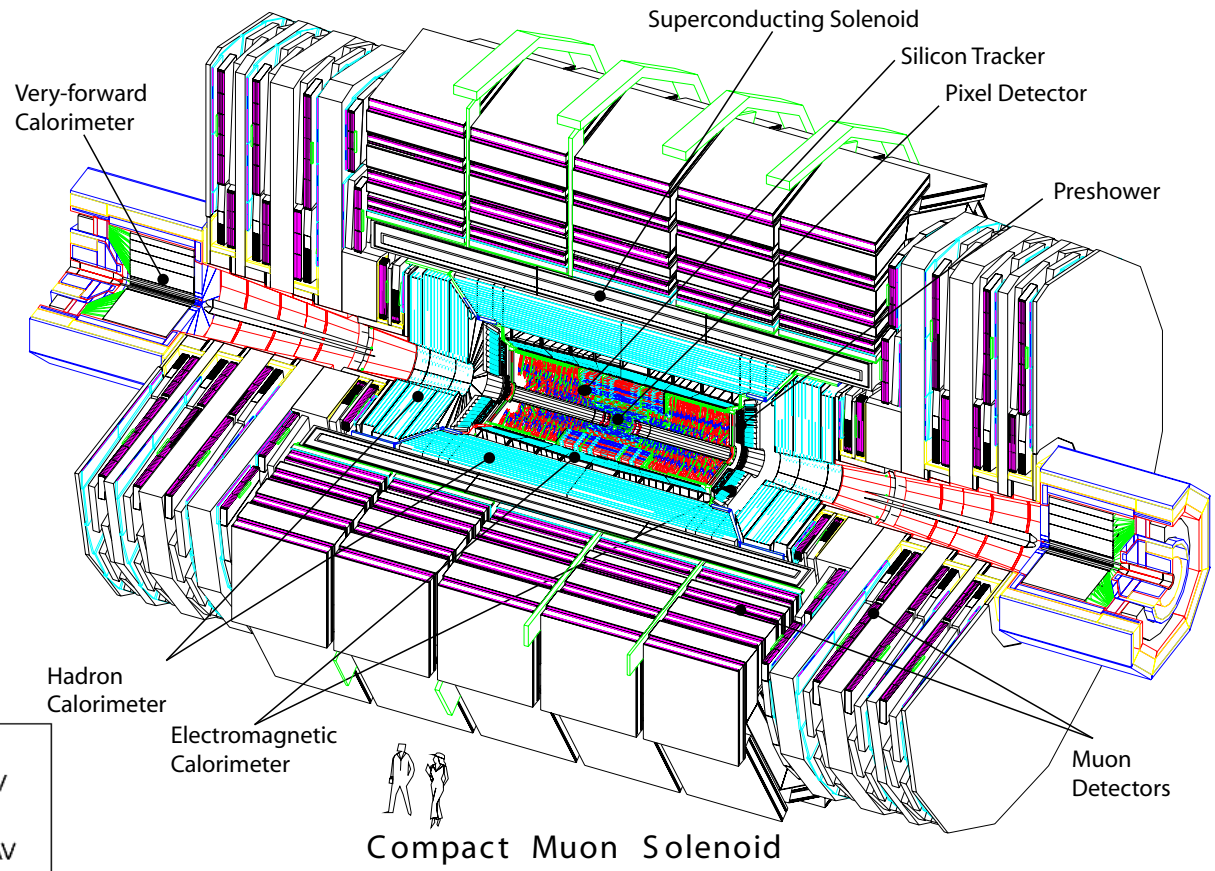
The CMS detector

- Design prioritization
  - lepton/photon ID
  - b/τ tagging
  - jets and $\not{E}_T$

Weight	12'500 t
Length	21.6 m
Diameter	15 m
Magnetic field	3.8 T

- Pixel detector
  - balanced resolution



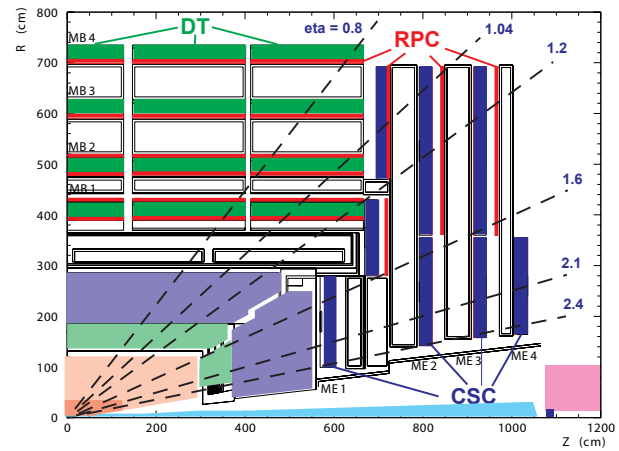


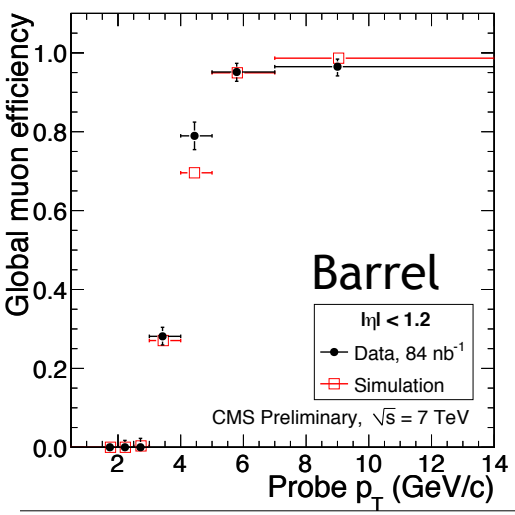
Component	Characteristics	resolutions
Pixel	3/2 Si layers	$\delta_z \approx 20\,\mu$m, $\delta_\phi \approx 10\,\mu$m
Tracker	10/12 Si strips	$\delta(p_\perp)/p_\perp \approx 1\%$
ECAL	$PbWO_4$	$\delta E/E \approx 3\%/\sqrt{E} \oplus 0.5\%$
HCAL (B)	Brass/Sc, $> 7.2\lambda$	$\delta E/E \approx 100\sqrt{E}\%$
HCAL (F)	Fe/Quartz	$\delta(\not{E}_T) \approx 0.98\sqrt{\sum E_T}$
Magnet	3.8 T solenoid	
Muons	DT/CSC + RPC	$\delta(p_\perp)/p_\perp \approx 10\%$ (STA)

JINST, 7, P10002

Muon reconstruction and identification

- Large muon acceptance $|\eta| < 2.4$
  - drift tubes
  - cathode strip chambers
  - resistive plate chambers
- Muon reconstruction/identification
  - global muon: outside-in reconstruction
  - tight muon: quality criteria against fakes
  - BDT: reduce fakes by another 50%
    - track 'kinks'
    - inner-outer matching
    - muon detector information

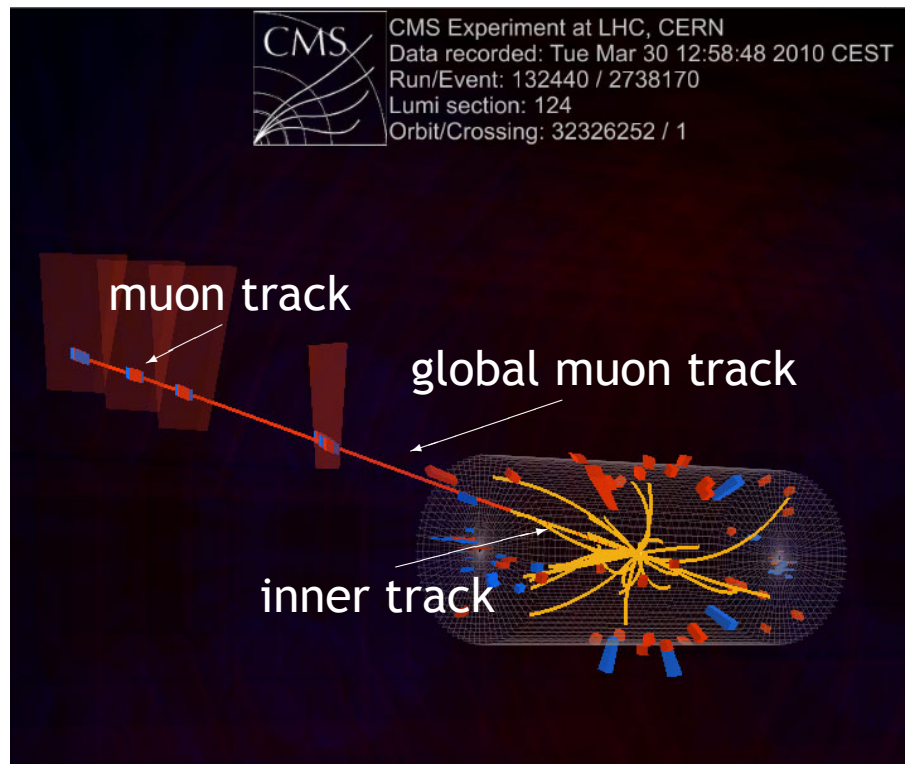




Muon misidentification

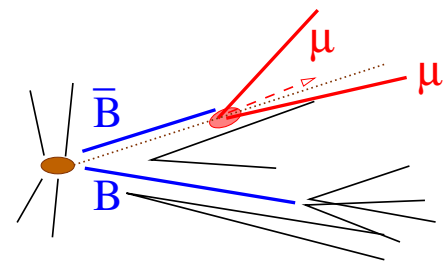
$$\varepsilon(\mu|\pi) \leq 0.15\%$$
$$\varepsilon(\mu|K) \leq 0.20\%$$
$$\varepsilon(\mu|p) \leq 0.10\%$$

measured in data and MC
(50% syst. uncertainty)

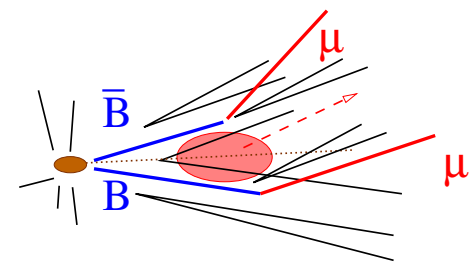


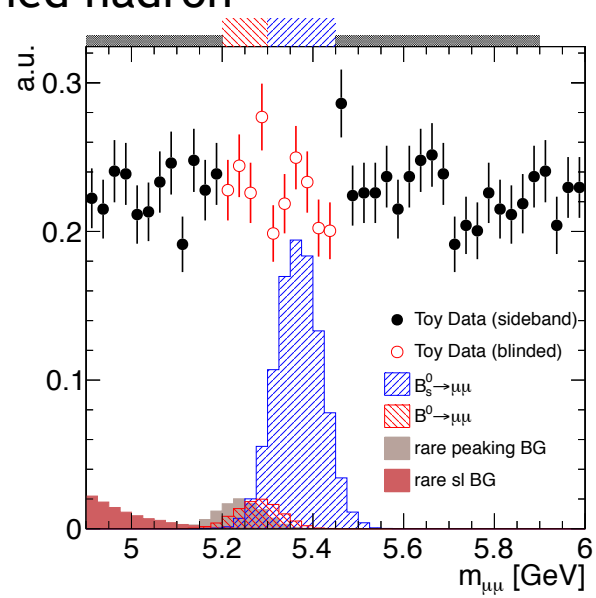
Analysis overview

- Signal $B_s^0 \to \mu^+\mu^-$
  - ▷ two muons from one decay vertex
    - well reconstructed secondary vertex
    - momentum aligned with flight direction
    - mass around $m_{B_s^0}$
    - isolated



- Background
  - ▷ combinatorial (from sidebands)
    - two semileptonic (B) decays (gluon splitting)
    - one semileptonic (B) decay and one misidentified hadron
  - ▷ rare single B decays (from MC simulation)
    - non-peaking, e.g. $B_s^0 \to K^-\mu^+\nu$, $\Lambda_b \to p\mu^+\nu$
    - peaking, e.g. $B_s^0 \to K^+K^-$





⇒ Critical issues
- ▷ optimized selection
- ▷ muon misidentification probability
  - enters quadratically for peaking bg
- ▷ pileup (isolation)

Methodology

- Measurement of $B_s^0 \to \mu^+\mu^-$ relative to normalization channel:
  - ▷ $B^\pm \to J/\psi\, K^\pm$, with well-known branching fraction
  - ▷ (nearly) identical selection to reduce systematic uncertainties

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = \frac{n^{\rm obs}_{B_s^0}}{\varepsilon_{B_s^0}\, N_{B_s^0}} = \frac{n^{\rm obs}_{B_s^0}}{\varepsilon_{B_s^0}\, \mathcal{L}\, \sigma(pp \to B_s^0)}$$

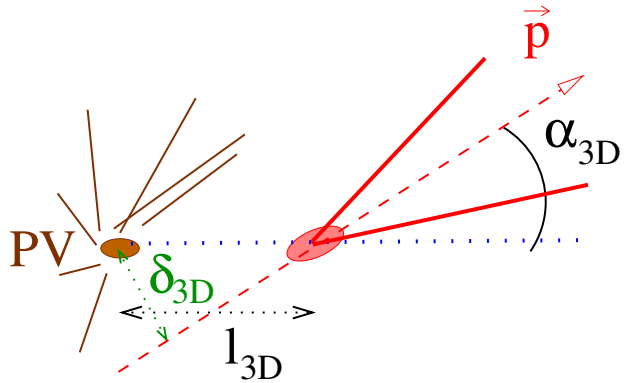
$$= \frac{n^{\rm obs}_{B_s^0}}{N(B^\pm \to J/\psi\, K^\pm)} \frac{A_{B^+}}{A_{B_s^0}} \frac{\varepsilon^{ana}_{B^+}}{\varepsilon^{ana}_{B_s^0}} \frac{\varepsilon^{\mu}_{B^+}}{\varepsilon^{\mu}_{B_s^0}} \frac{\varepsilon^{trig}_{B^+}}{\varepsilon^{trig}_{B_s^0}} \frac{f_u}{f_s}\, \mathcal{B}(B^+ \to J/\psi\,[\mu^+\mu^-]K)$$

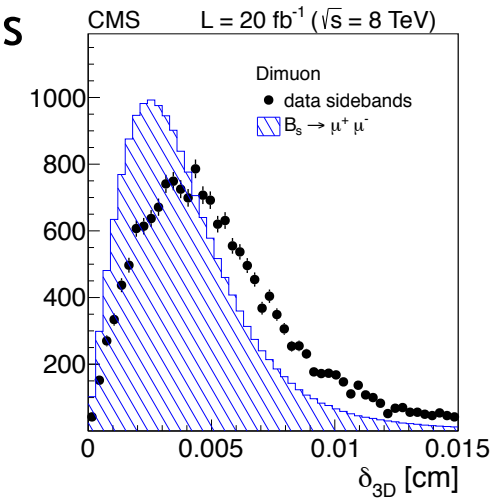
- Calibration of MC with reconstructed exclusive decays
  - ▷ $B^\pm \to J/\psi\, K^\pm$: normalization with high statistics
  - ▷ $B_s^0 \to J/\psi\, \phi$: B_s^0 signal MC ($p_\perp$ and isolation)
- Two 'channels' per dataset
  - ▷ barrel: both muons with $|\eta| < 1.4$
  - ▷ endcap: 1-2 muon(s) with $|\eta| > 1.4$

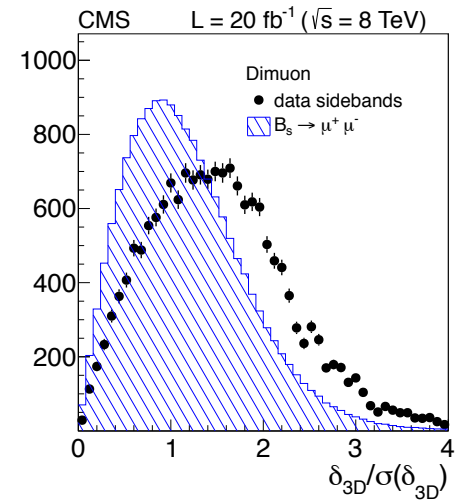
$5\,\text{fb}^{-1}$ at $\sqrt{s} = 7\,$TeV taken in 2011
$20\,\text{fb}^{-1}$ at $\sqrt{s} = 8\,$TeV taken in 2012

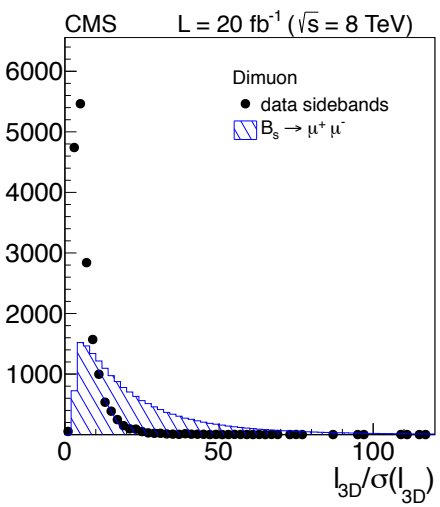
Discriminating variables

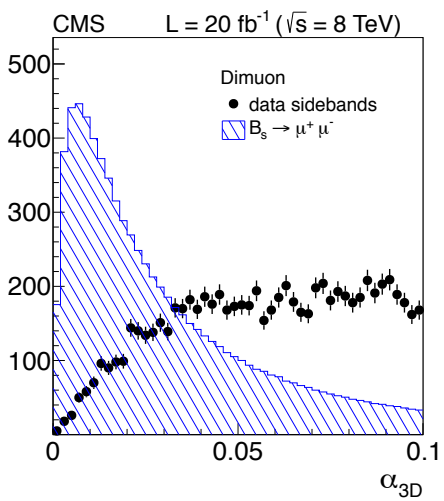
- Vertexing
  - ▷ primary vertex w/o the two muons
  - ▷ secondary vertex of the two muons

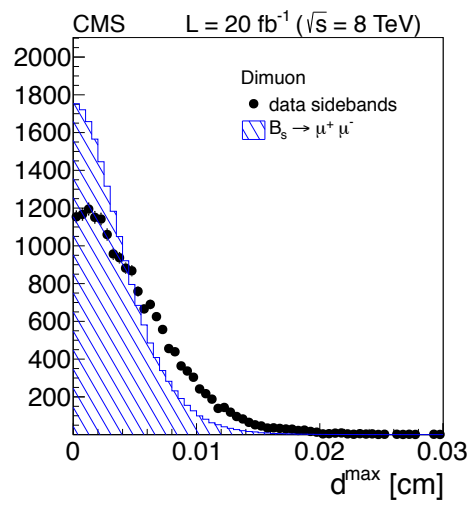


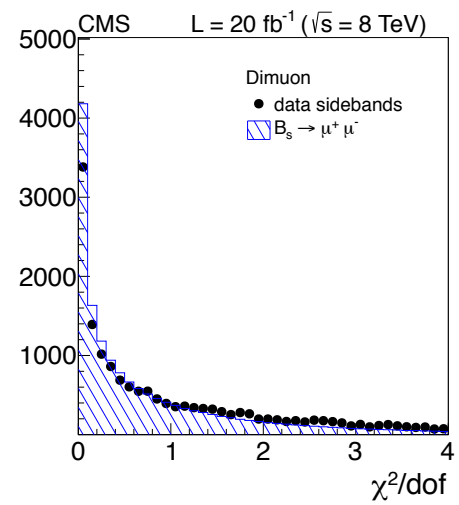












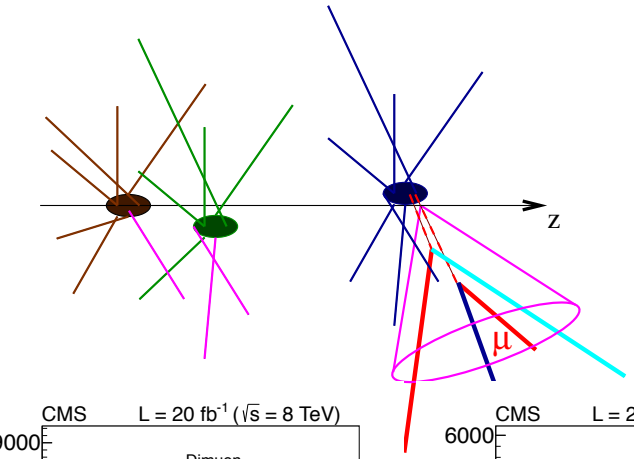
Isolation Variables

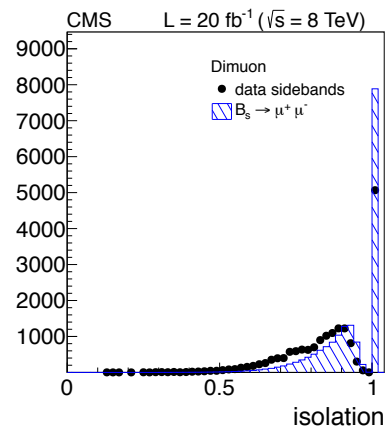
- Relative isolation of dimuon

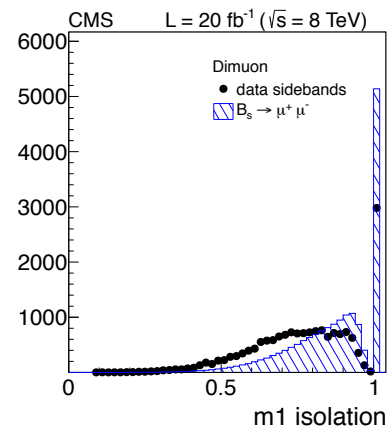
$$I = \frac{p_\perp(\mu^+\mu^-)}{p_\perp(\mu^+\mu^-) + \sum\limits_{\Delta R<1} p_\perp}$$

  - ▷ in cone around dimuon momentum
  - ▷ for tracks in cone with $\Delta R < 0.7$
    - with $p_\perp > 0.9$ GeV
    - either associated to same PV as candidate
    - or with $d_{ca} < 500\,\mu$m

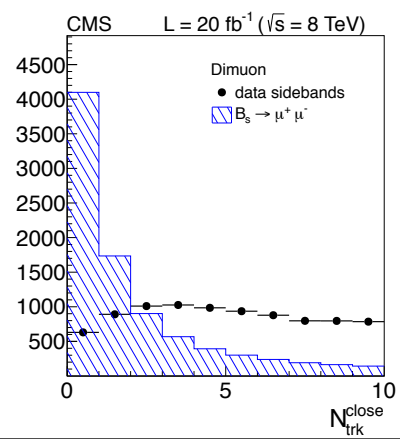
      (d_{ca} = distance of closest approach)

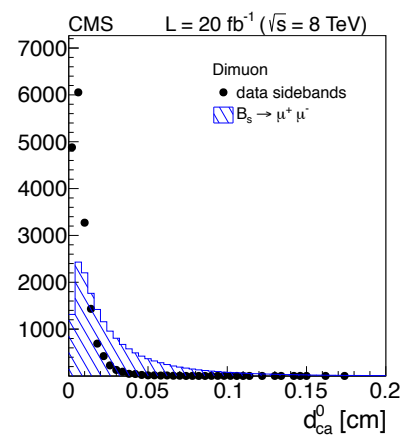






- Muon isolation
  - ▷ $\Delta R < 0.5$, $p_\perp > 0.5$ GeV and $d_{ca} < 1$ mm
- Number of tracks close to SV
  - ▷ $p_\perp > 0.5$ GeV and $d_{ca} < 300\,\mu$m
- Closest track to SV
  - ▷ d^0_{ca}

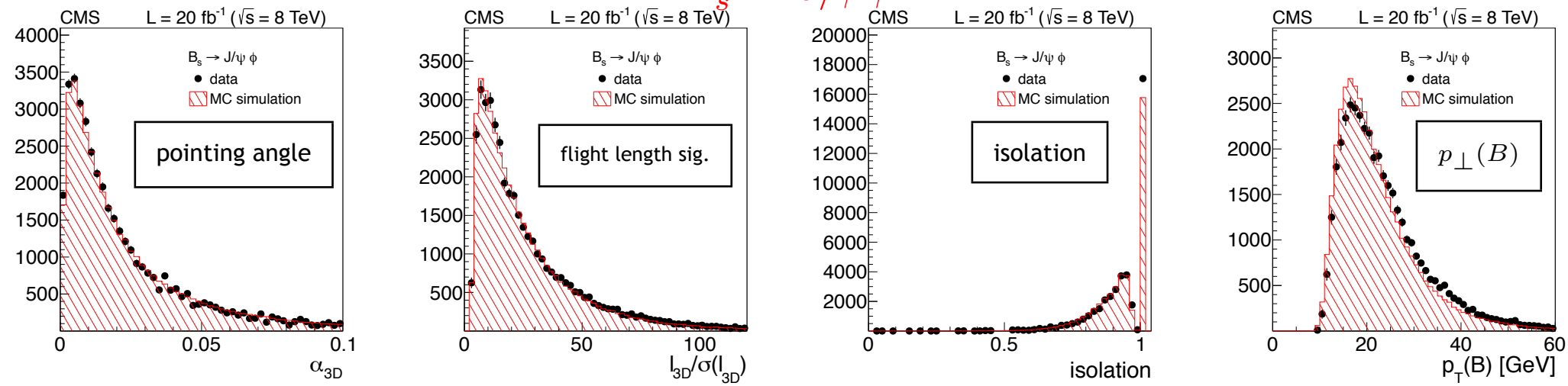




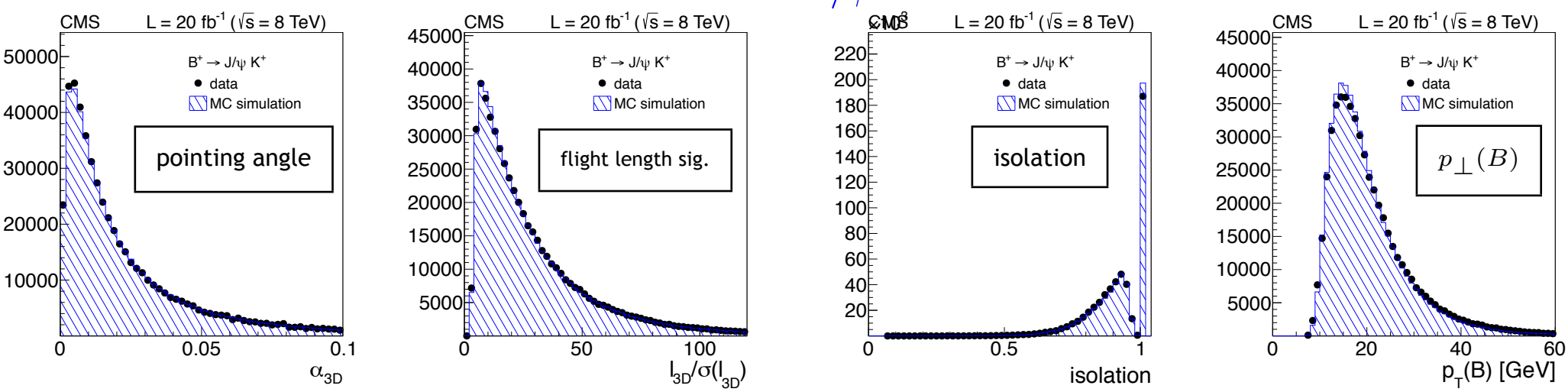
MC simulation vs. data

- Comparison of sideband-subtracted distributions
  - ▷ in general good agreement

$B_s^0 \to J/\psi\,\phi$



$B^\pm \to J/\psi\,K^\pm$



MVA selection

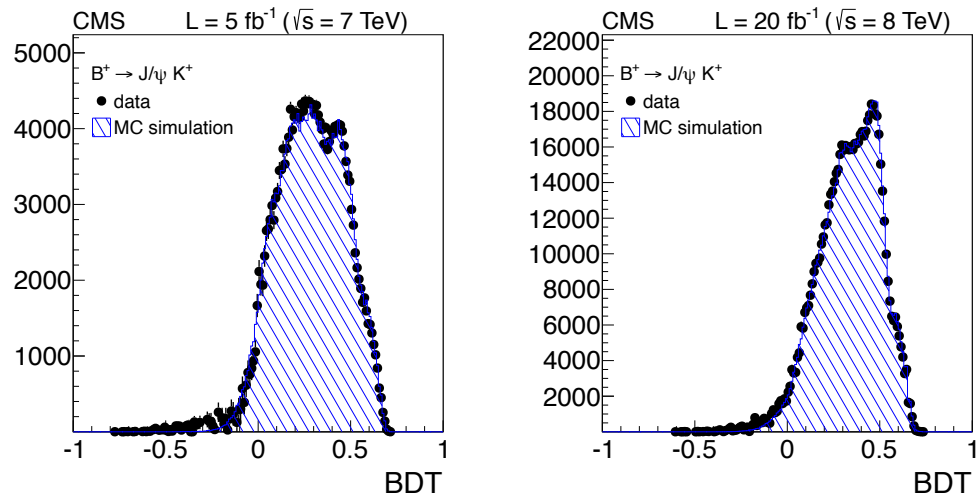
- BDT training
  - ▷ TMVA framework
  - ▷ signal: $B_s^0 \to \mu^+\mu^-$ MC simulation
  - ▷ background: dimuon data mass sidebands
  - ▷ avoid selection bias
    - split data randomly into three subsets (0,1,2)
    - train on 1, test on 2, apply on 3. etc.
  - → in each channel, have 3 BDTs
- Studies
  - ▷ selection efficiency in high and low mass sidebands
  - ▷ signal MC with shifted mass
  - ▷ pileup for dimuons (and normalization/control samples)
- Selection of normalization and control samples with identical BDTs
  - ▷ slightly modified variables (e.g. dimuon vertex fit quality)
  - ▷ isolation variables: ignore hadronic particles from B decay

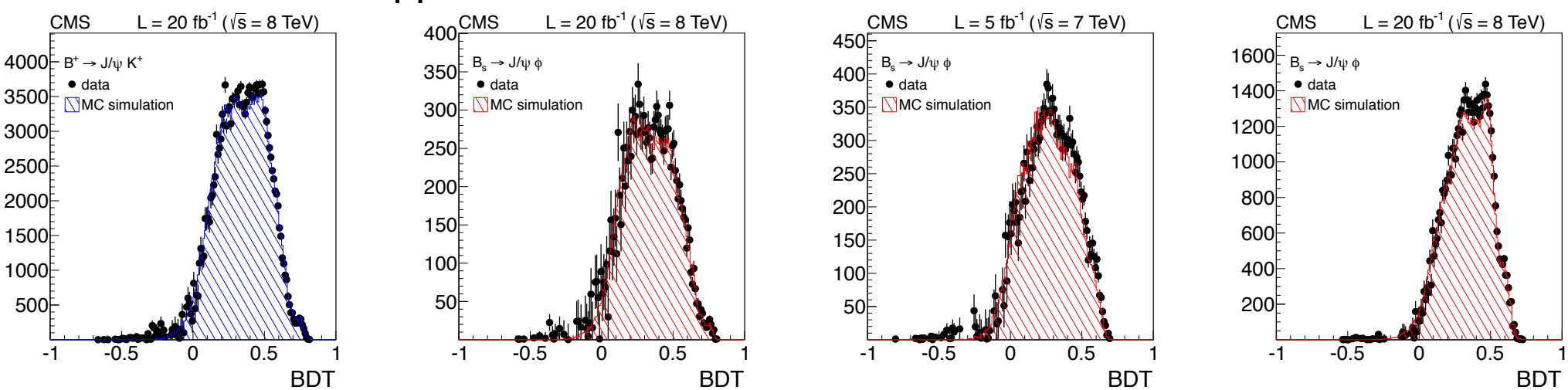
MC simulation vs. data (II)

- Differences between data and MC used as systematic uncertainties
  - $B^{\pm} \to J/\psi\, K^{\pm}$: <3.0%
  - $B_s^0 \to J/\psi\, \phi$: <9.5% (2011)
    <3.5% (2012)

    used for signal ε uncertainty

- Mass scale uncertainty
  - ψ and Υ to dimuon decays
    barrel: < 6 MeV
    endcap: < 7 MeV
  - correction applied

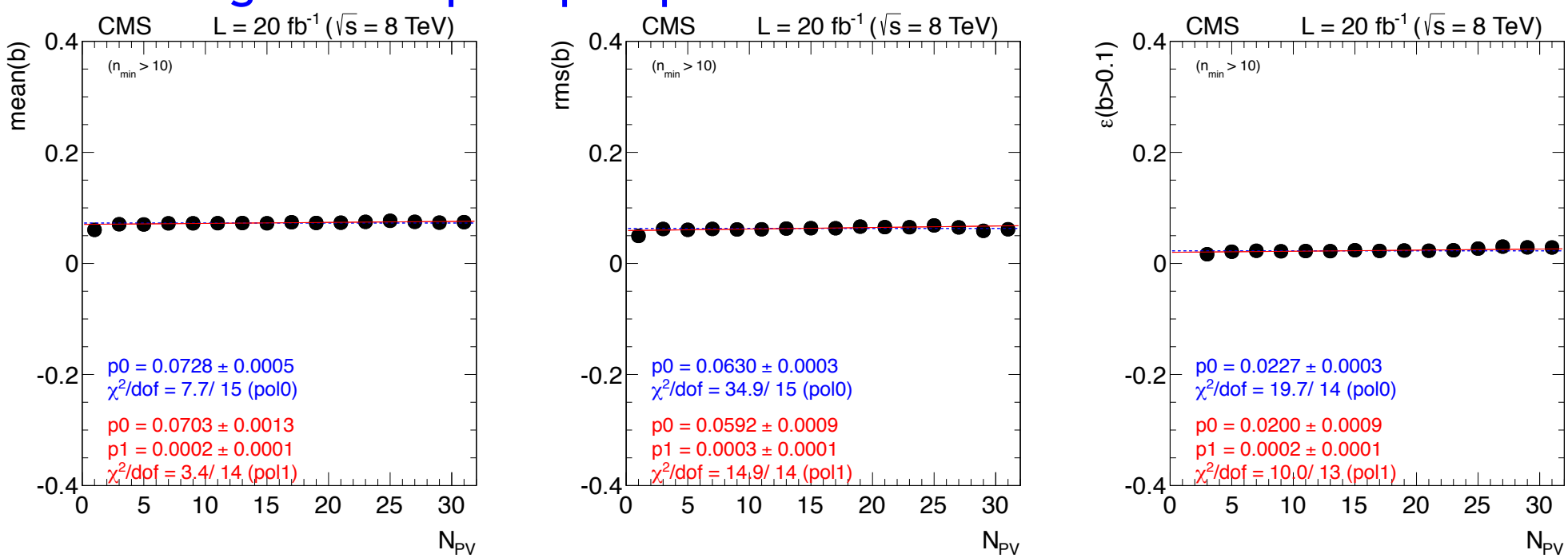




Pileup independence

- Average number of interactions per bunch crossing
  - ▷ 2011: ≈ 9, 2012: ≈ 21
- Pileup independence checked
  - ▷ Signal MC event samples with pileup
    - every single variable used in BDT is shown to be pileup independent
  - ▷ Data studies with BDT output distribution vs. N_{PV}
    - ▷ mean and RMS, efficiency of BDT requirement

$\Rightarrow$ No significant pileup dependence observed



BDT Categories

- Two approaches for interpretation
  - ▷ 1D-BDT
    - (optimized) cut on BDT output

$b >$	barrel	endcap
2011	0.29	0.29
2012	0.38	0.39

independent data set used

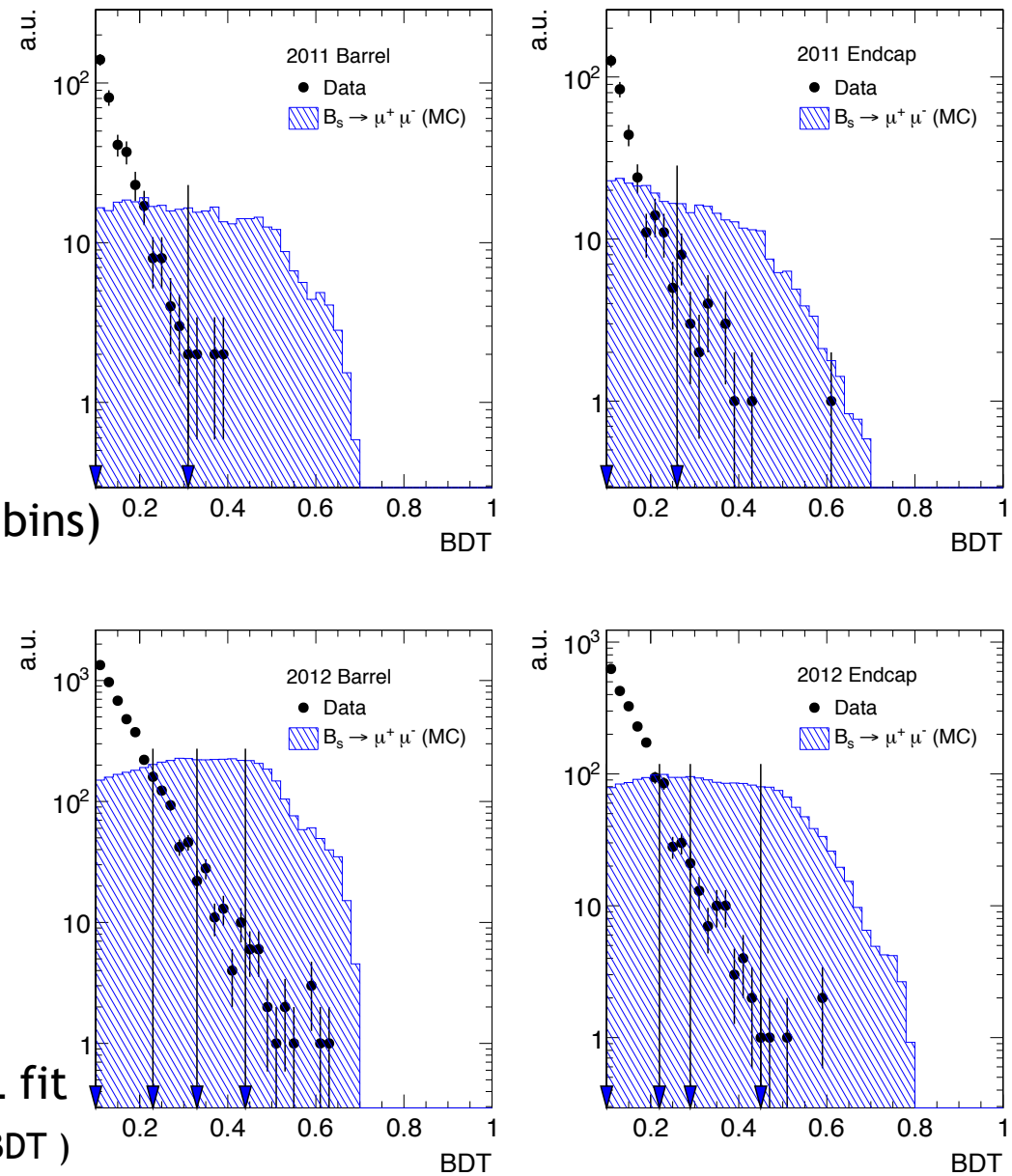
→ 4 mass distributions

  - ▷ categorized-BDT
    - per channel 2-4 categories (BDT bins)

min. bin edges	1	2	3	4
2011 barrel	0.10	0.31	-	-
2011 endcap	0.10	0.29	-	-
2012 barrel	0.10	0.23	0.33	0.44
2012 endcap	0.10	0.22	0.29	0.45

equalized expected signal yield

→ 12 mass distributions

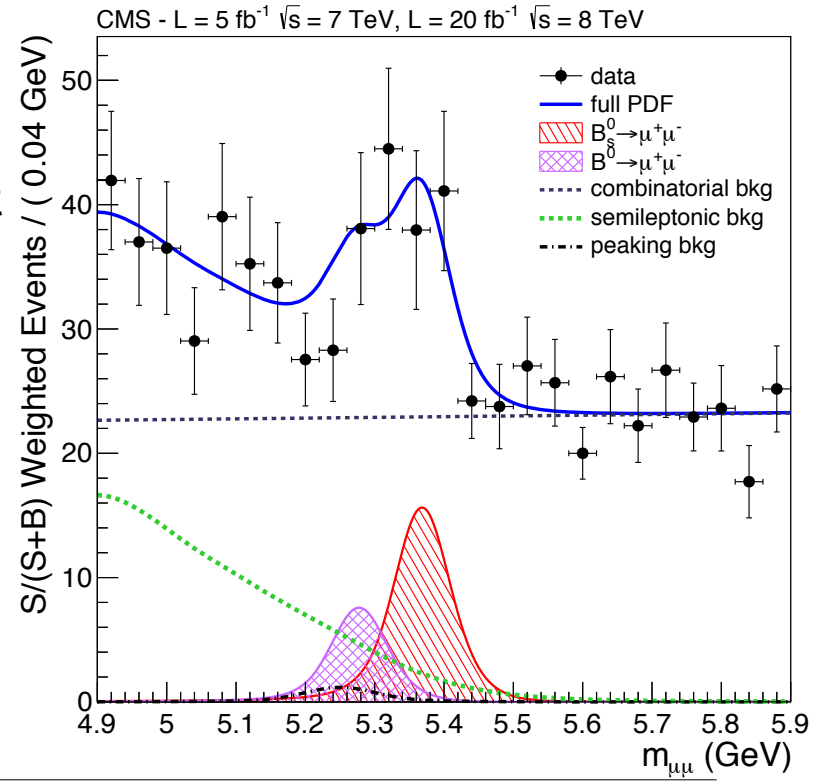


- Strategy (decided before unblinding)
  - ▷ $B^0 \to \mu^+\mu^-$ 1D-BDT: UL with CL_S
  - ▷ $B_s^0 \to \mu^+\mu^-$ categorized-BDT: UML fit

(best expected sensitivity with categorized-BDT)

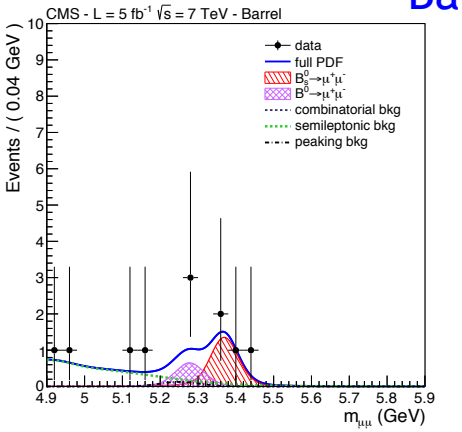
Unbinned maximum likelihood fit

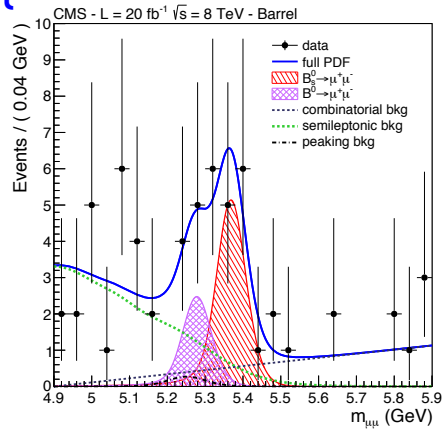
- Probability distribution functions
  - ▷ peaking components: Crystal Ball (w/ and w/o Gaussian)
  - ▷ combinatorial background: polynomial of first degree
  - ▷ $b \to u\mu\bar{\nu}$ background: Gaussian kernels for MC-predicted mixture
  - ▷ per-event mass resolution included (excellent data/MC agreement)
- Fit for B_s^0 and B^0 simultaneously
  - ▷ peaking background
    - constrained to expectation
    - normalized to measured B^+ yield
    - yield cross checked on independent data set
  - ▷ semileptonic background
    - fixed shape
    - floating normalization within uncertainties
    - (dominated by unknown $\Lambda_b \to p\mu\nu$)
  - ▷ combinatorial background
    - no constraint on slope
    - validated with independent data set
    - varied functional form



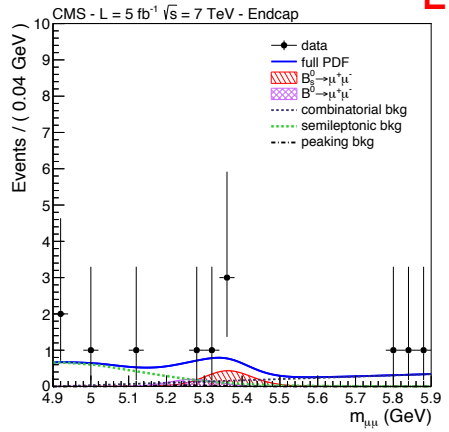
Expectations and observation (1D-BDT)

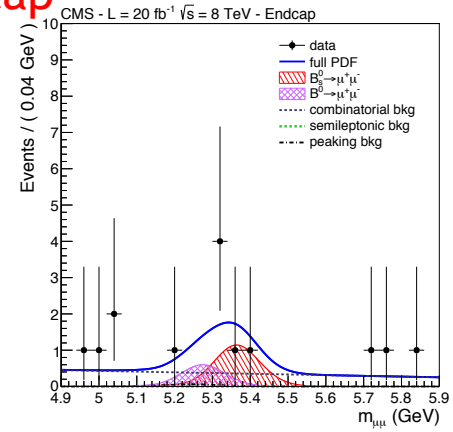
Barrel





Endcap





- Summary numbers for 1D-BDT approach in B_s^0 and B^0 signal regions
  - ▷ signal regions

B^0 : $5.20 < m < 5.30$ GeV

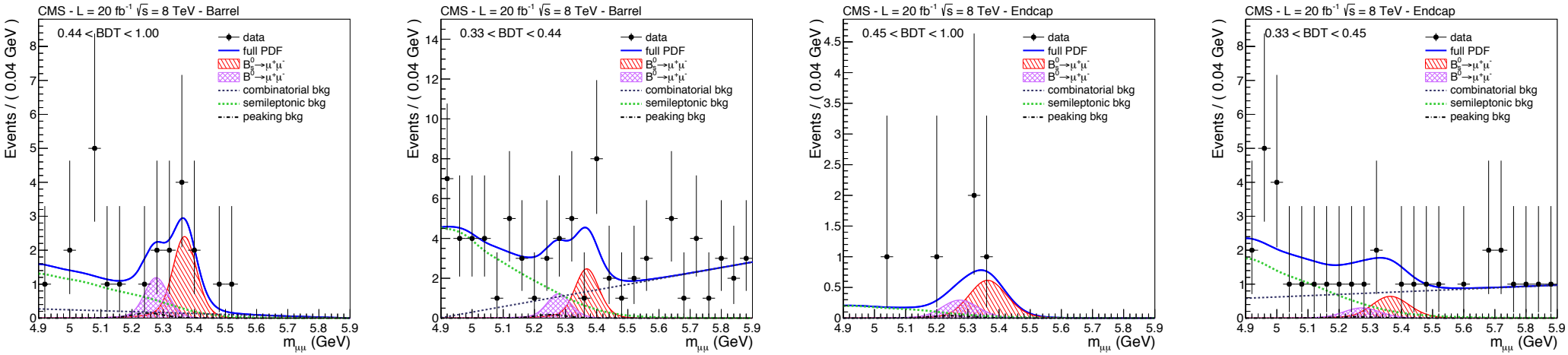
B_s^0 : $5.30 < m < 5.45$ GeV

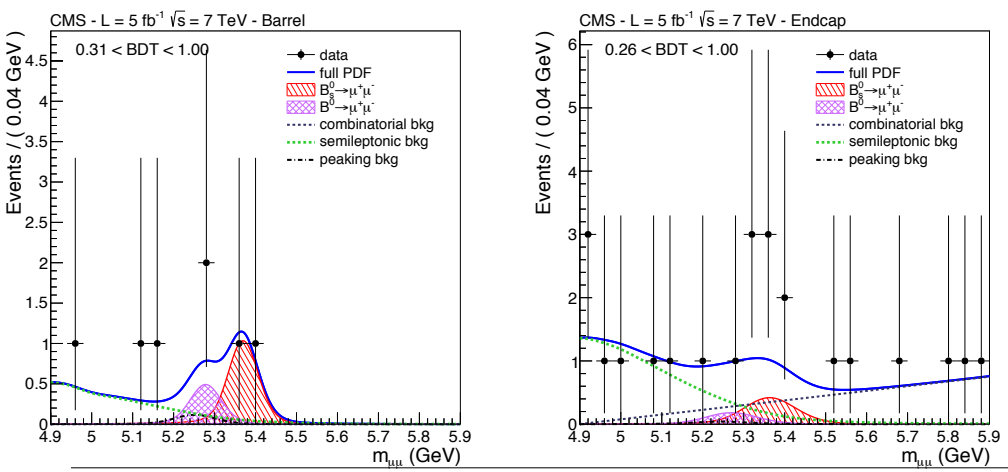
	2011 barrel		2012 barrel	
	$B^0 \to \mu^+\mu^-$	$B_s^0 \to \mu^+\mu^-$	$B^0 \to \mu^+\mu^-$	$B_s^0 \to \mu^+\mu^-$
$\varepsilon_{\text{tot}}[\%]$	0.33 ± 0.03	0.30 ± 0.04	0.24 ± 0.02	0.23 ± 0.03
$N_{\text{signal}}^{\text{exp}}$	0.27 ± 0.03	2.97 ± 0.44	1.00 ± 0.10	11.46 ± 1.72
$N_{\text{total}}^{\text{exp}}$	1.3 ± 0.8	3.6 ± 0.6	7.9 ± 3.0	17.9 ± 2.8
N_{obs}	3	4	11	16

	2011 endcap		2012 endcap	
	$B^0 \to \mu^+\mu^-$	$B_s^0 \to \mu^+\mu^-$	$B^0 \to \mu^+\mu^-$	$B_s^0 \to \mu^+\mu^-$
$\varepsilon_{\text{tot}}[\%]$	0.20 ± 0.02	0.20 ± 0.02	0.10 ± 0.01	0.09 ± 0.01
$N_{\text{signal}}^{\text{exp}}$	0.11 ± 0.01	1.28 ± 0.19	0.30 ± 0.03	3.56 ± 0.53
$N_{\text{total}}^{\text{exp}}$	1.5 ± 0.6	2.6 ± 0.5	2.2 ± 0.8	5.1 ± 0.7
N_{obs}	1	4	3	4

Unbinned maximum likelihood fit (II)

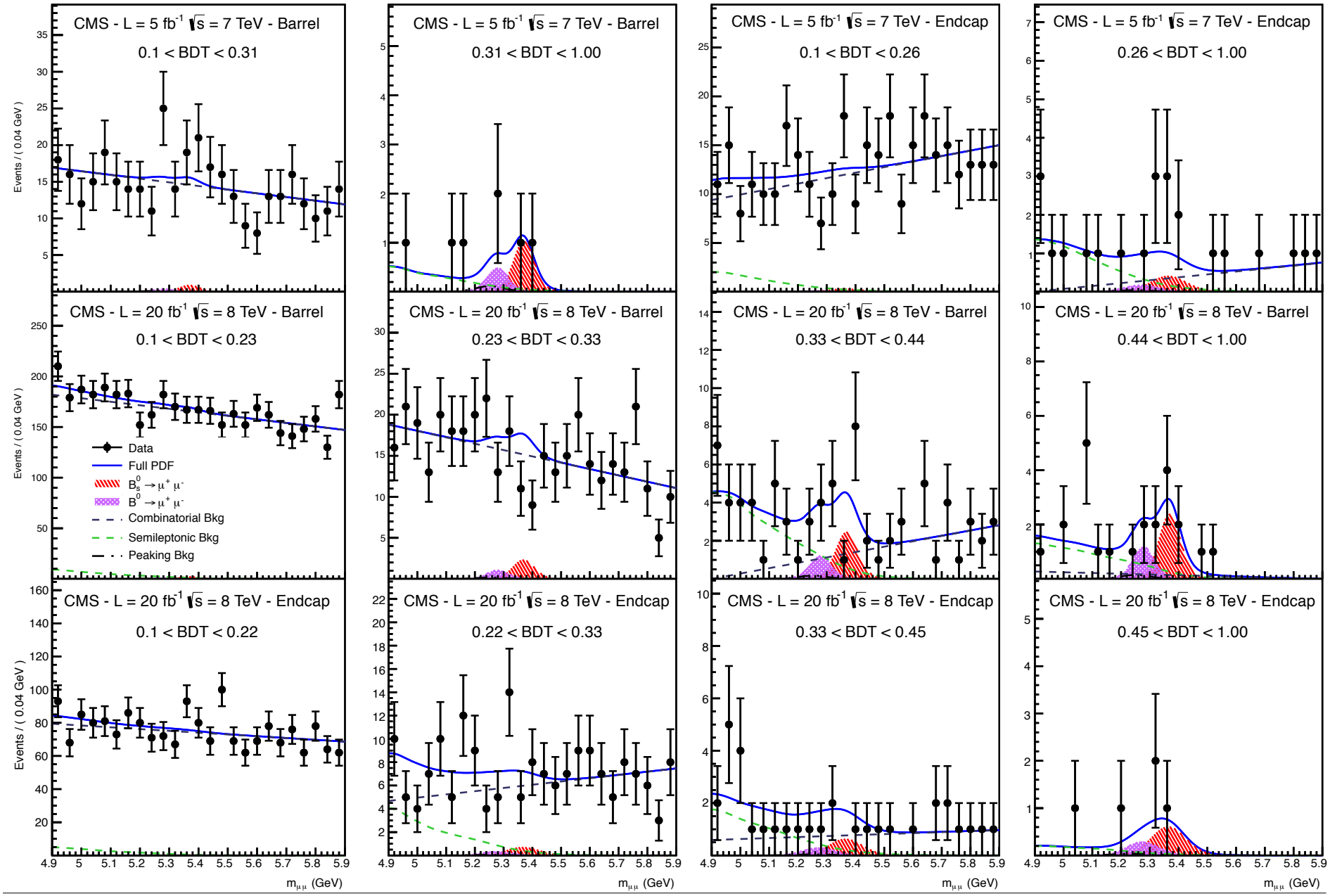
- Illustration of the UML fits
  - ▷ highest (2nd highest) S/B categories for (barrel, endcap) $\times$ (2011, 2012)





- Signal candidate distributions
  - ▷ consistent with (signal!) expectations
    - kinematic variables
    - vertexing variables
    - number of pp interactions

All BDT bins


CMS - L = 5 fb^{-1} $\sqrt{s}$ = 7 TeV - Barrel
0.1 < BDT < 0.31
CMS - L = 5 fb^{-1} $\sqrt{s}$ = 7 TeV - Barrel
0.31 < BDT < 1.00
CMS - L = 5 fb^{-1} $\sqrt{s}$ = 7 TeV - Endcap
0.1 < BDT < 0.26
CMS - L = 5 fb^{-1} $\sqrt{s}$ = 7 TeV - Endcap
0.26 < BDT < 1.00
CMS - L = 20 fb^{-1} $\sqrt{s}$ = 8 TeV - Barrel
0.1 < BDT < 0.23
CMS - L = 20 fb^{-1} $\sqrt{s}$ = 8 TeV - Barrel
0.23 < BDT < 0.33
CMS - L = 20 fb^{-1} $\sqrt{s}$ = 8 TeV - Barrel
0.33 < BDT < 0.44
CMS - L = 20 fb^{-1} $\sqrt{s}$ = 8 TeV - Barrel
0.44 < BDT < 1.00
CMS - L = 20 fb^{-1} $\sqrt{s}$ = 8 TeV - Endcap
0.1 < BDT < 0.22
CMS - L = 20 fb^{-1} $\sqrt{s}$ = 8 TeV - Endcap
0.22 < BDT < 0.33
CMS - L = 20 fb^{-1} $\sqrt{s}$ = 8 TeV - Endcap
0.33 < BDT < 0.45
CMS - L = 20 fb^{-1} $\sqrt{s}$ = 8 TeV - Endcap
0.45 < BDT < 1.00
Events / (0.04 GeV)
Data
Full PDF
$B_s^0 \to \mu^+\mu^-$
$B^0 \to \mu^+\mu^-$
Combinatorial Bkg
Semileptonic Bkg
Peaking Bkg
$m_{\mu\mu}$ (GeV)

Results

- Results of the UML fit in the categorized-BDT approach

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (3.0^{+1.0}_{-0.9}) \times 10^{-9}$$
$$\mathcal{B}(B^0 \to \mu^+\mu^-) = (3.5^{+2.1}_{-1.8}) \times 10^{-10}$$

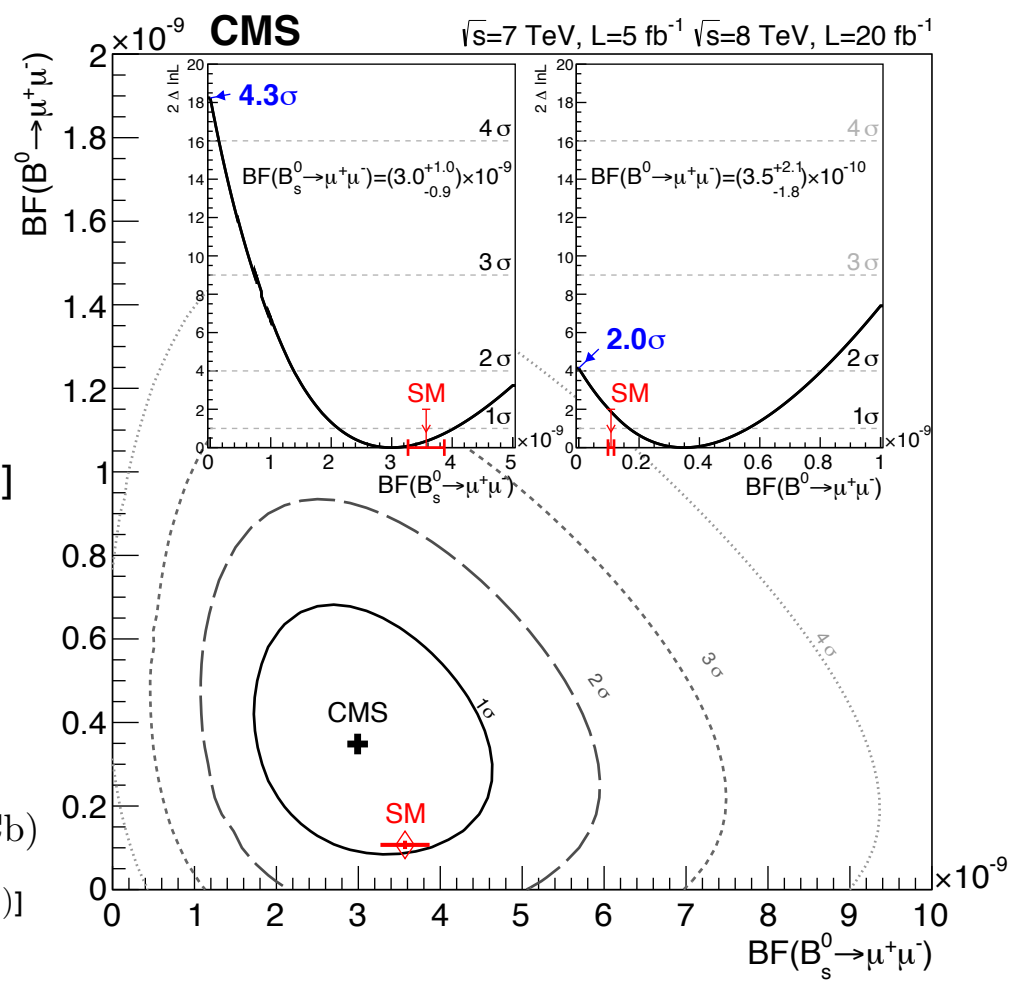
$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = [3.0^{+0.9}_{-0.8}(\text{stat})^{+0.6}_{-0.4}(\text{syst})] \times 10^{-9}$

- ▷ significances
  - $B_s^0 \to \mu^+\mu^-$: 4.3σ [expected 4.8σ (median)]
  - $B^0 \to \mu^+\mu^-$: 2.0σ

  [B^0 (B_s^0) left floating for B_s^0 (B^0) significance]
- ▷ UML fit in the 1D-BDT approach
  - $B_s^0 \to \mu^+\mu^-$: 4.8σ [expected 4.7σ (median)]
- ▷ external input

$$\mathcal{B}(B^+) = (6.0 \pm 0.2) \times 10^{-5} \text{ (PDG)}$$
$$f_s/f_u = 0.256 \pm 0.020 \pm 0.013_{\text{extrapol.}} \text{ (LHCb)}$$

[cross check with $B_s^0 \to J/\psi\,\phi$ and $B^\pm \to J/\psi\,K^\pm$ vs. $(p_\perp, \eta)$]



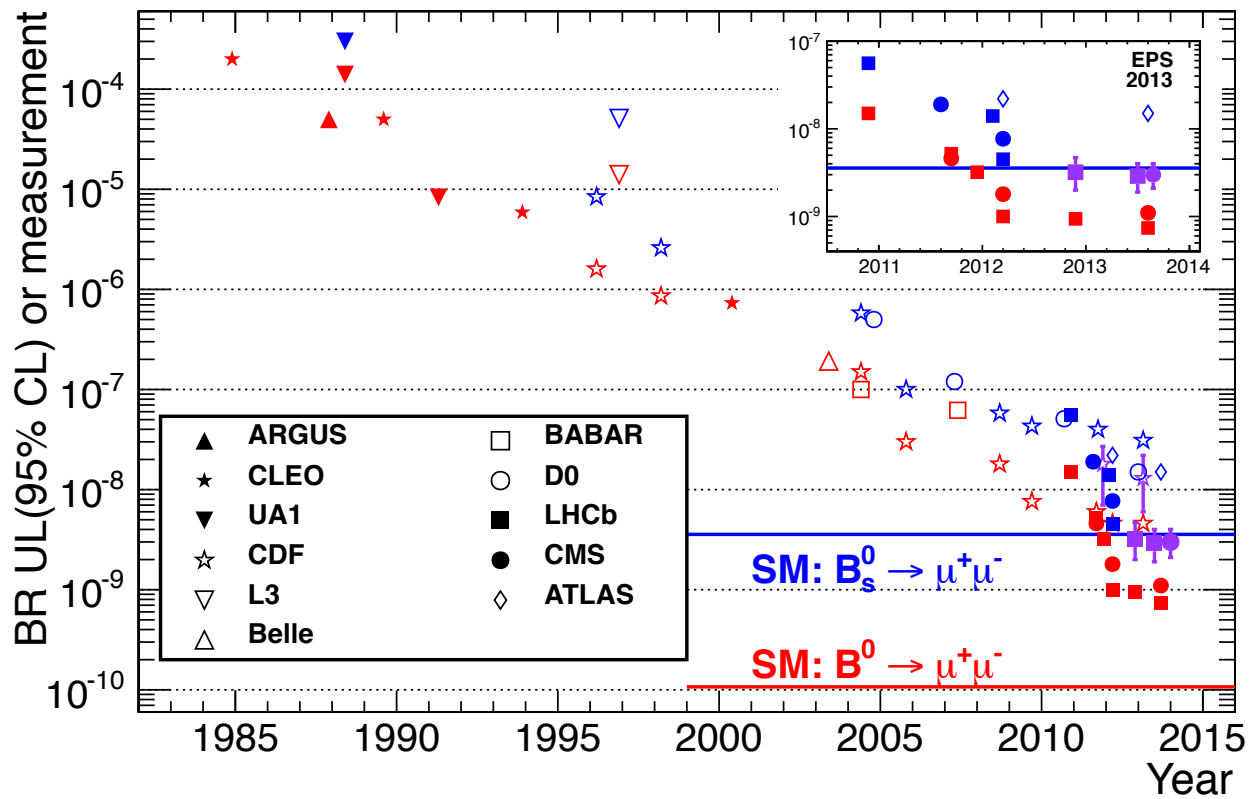
CL_S for the 1D-BDT approach

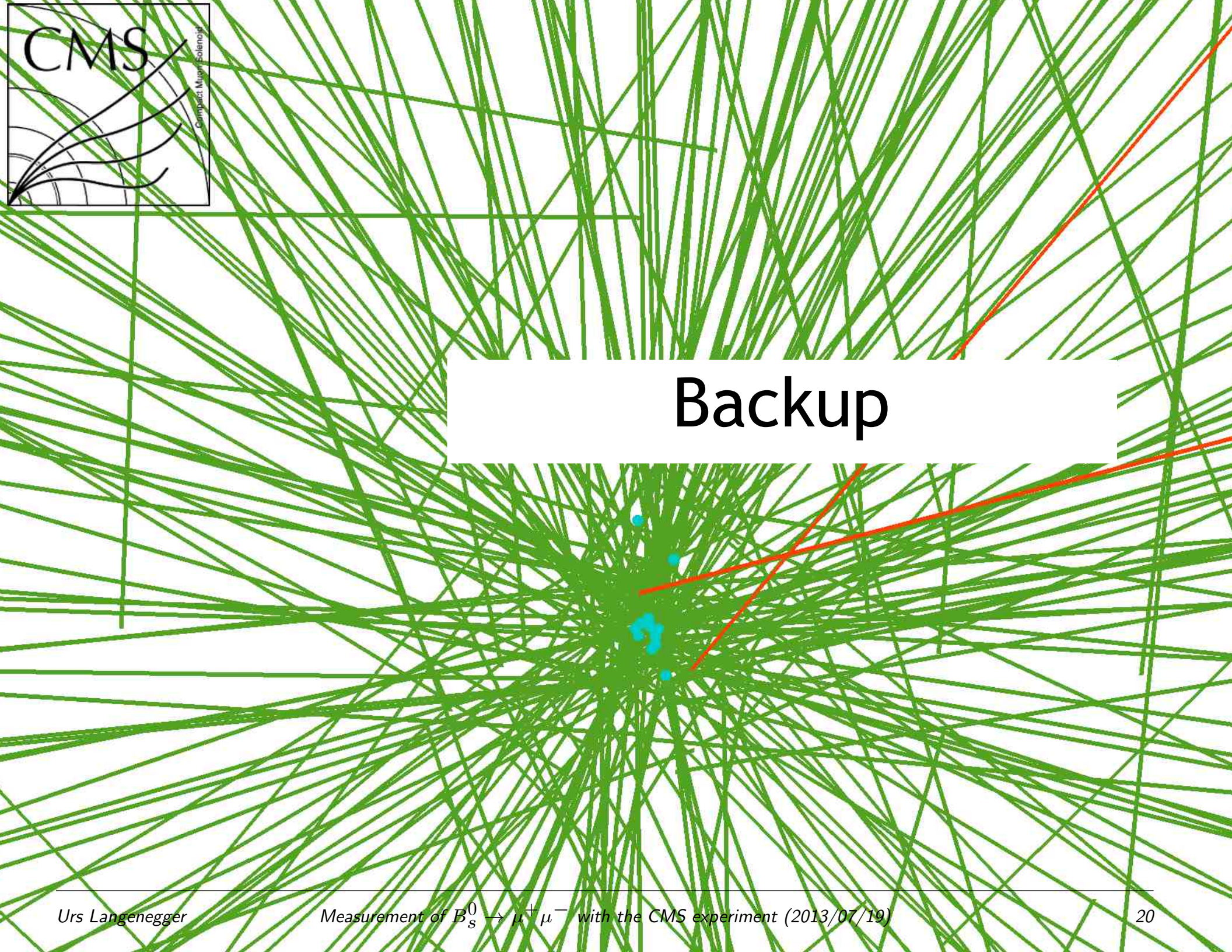
- Upper limit on $B^0 \to \mu^+\mu^-$

$$\mathcal{B}(B^0 \to \mu^+\mu^-) < 1.1 \times 10^{-9} \quad 95\%\text{CL}$$

Conclusions

- Measurement of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ and upper limit on $\mathcal{B}(B^0 \to \mu^+\mu^-)$
  - ▷ substantial improvements to previous analysis
    - muon identification with BDT
    - analysis selection with BDT
    - UML fit to mass distributions
  - ▷ 4.3σ significance of observation (4.8σ with 1D-BDT approach)
  - ▷ consistent with SM
- The future: $B^0 \to \mu^+\mu^-$
  - ▷ 'parked' data
    had. two-body B decays
  - ▷ Run 2
    up to 300 fb^{-1}
  - ▷ $B_s^0 \to \mu^+\mu^-$
    as 'calibration'

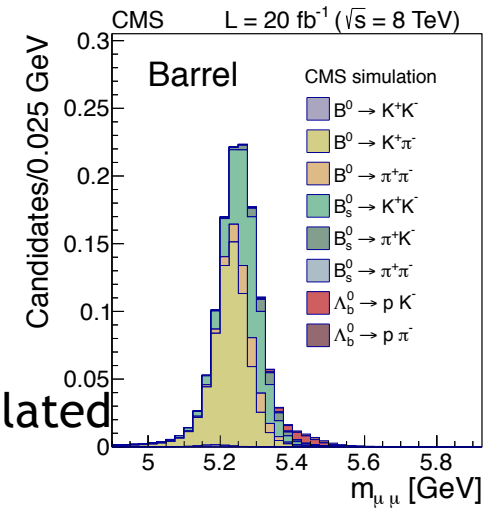


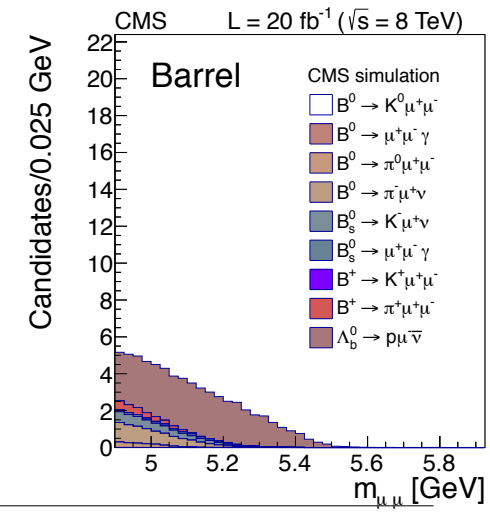


Backup

Systematics

- Hadronization probability ratio f_s/f_u from LHCb [JHEP 04, 001 (2013)]
  - additional 5% systematics for possible $p_\perp$ or η dependence
  - in-situ studies show no $p_\perp$ dependence
    ratio of $B^\pm \to J/\psi\, K^\pm$ vs $B_s^0 \to J/\psi\, \phi$
- Rare decays
  - hadron to muon misidentificaton probability
    $K_S^0 \to \pi^+\pi^-$, $\Lambda \to p\pi$, and $D^{*+} \to D^0(K^-\pi^+)\pi^+$
    50% uncertainty, treating pions/kaons/protons as uncorrelated
  - branching fraction uncertainties
  - $\Lambda_b \to p\mu\bar{\nu}$:
    large range of predictions in literature
    take average (6.5×10^{-4}) and assign 100% uncertainty
    (note that invariant mass covers B_s^0 signal region)
- Normalization
  - 5% from yield fits





Trigger: $B_s^0 \to \mu^+\mu^-$ and $B^\pm \to J/\psi\, K^\pm$

- Dimuon trigger
  - ▷ L1 (hardware) trigger
    a few kHz at current peak luminosities
  - ▷ High-level trigger
    full tracking and vertexing
- HLT $B_s^0 \to \mu^+\mu^-$
  - ▷ two muons with opposite charge
  - ▷ inv. mass $4.8 < m_{\mu\mu} < 6.0$ GeV
  - ▷ $\mathcal{P}(\chi^2/dof) > 0.5\%$
- HLT $B^\pm \to J/\psi\, K^\pm$ and $B_s^0 \to J/\psi\, \phi$
  - ▷ two muons with opposite charge, $2.9 < m_{\mu\mu} < 3.3$ GeV
  - ▷ $\cos\alpha > 0.9$, $\mathcal{P}(\chi^2/dof) > 15\%$
  - → 'displaced' J/ψ

Trigger efficiency $40 - 80\%$
- ▷ after analysis selection
- ▷ time-dependence in MC

Determination
- ▷ MC simulation
- ▷ data
- → systematics from difference