The MPI Concept of Time-Dependent Fits at Belle II:

Vladimir Chekelian   Max Planck Institute for Physics, Munich
Implemantation on ALPOS: Daniel Britzger (MPI) and V.C.

The xFitter Workshop, Minsk, 18-20.03.2019

Outline:

1. *Introduction to the subject*
2. *Caveats of the traditional concept of the TD fit based on a convolution of physics with dt resolution function*
3. *Impact of the tiny beam spot size in Belle II on the dt resolution function*
4. *The MPI concept with pdfs determined on fly for every data event using re-weighted sample of MC events*
5. *Performance of the MPI approach in the TDCPV fit of B0 → Jpsi K0S*
6. *Conclusions & call for input and expertise*

Introduction: time-dependent fits & traditional approach

Time-dependent fits *are used to determine lifetimes (B0/B$^{\pm}$), B0 mixing parameters and CP violation parameters*

Traditional approach *used by Belle & BaBar:* *unbinned maximum likelihood fit to $\Delta t = t_{Bsig} - t_{Btag}$*

$$\text{maximum} \quad \boldsymbol{L} = \prod_i \boldsymbol{P}(\Delta \boldsymbol{t_i}, physics \;\&\; event\ reco\ parameters)$$

with $\boldsymbol{P}(\Delta \boldsymbol{t_i})$ *calculated as a convolution of theory and Δt resolution function:*

$$\boldsymbol{Psig}_{,\,\boldsymbol{bkg}}(\Delta \boldsymbol{t}) = \int_{-\infty}^{+\infty} \boldsymbol{d}(\Delta \boldsymbol{t}') \;\; \wp_{\mathbf{sig,\,bkg}}(\Delta \boldsymbol{t}') \;\; \boldsymbol{R}_{\mathbf{sig,\,bkg}}(\Delta \boldsymbol{t}\text{-}\Delta \boldsymbol{t}')$$

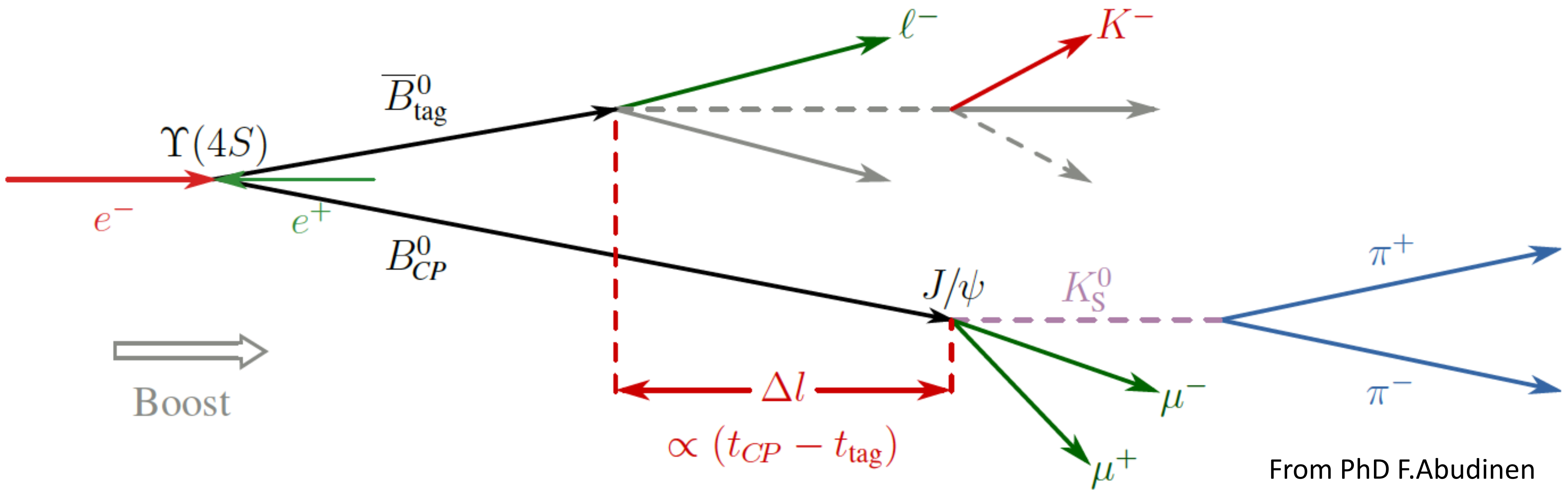
It is assumed that

- *dt resolution function is related entirely to detector effects and fully factorized from physics parameters of interest*
- *reconstructed efficiency is uniform in dt, dt_true*

Resolution function in Belle

- *complicated convolution of four components (with sub-classes) related to Bsig, to Btag with additional smearing due to tracks from charm and K0S, and to kinematic approximation that B0s are at rest in the Y(4S) frame*
- *it was a success, especially due invention of external estimators of quality such as "ξ"("h")*

Measurement of CP violating parameters



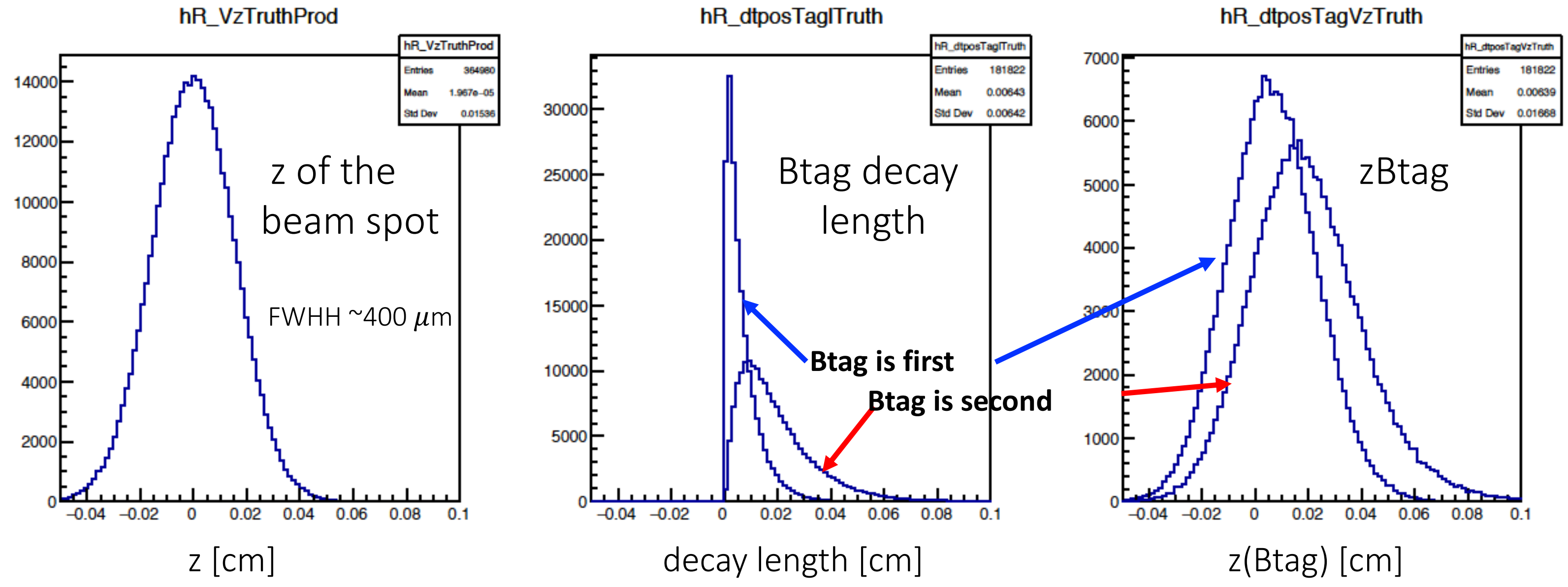
From PhD F.Abudinen

- *due to boost of Y(4S) one can measure distance between decays of Bsig and Btag and calculate dt=B0sig-B0tag*
- *CP violation & mixing parameters and B0 lifetime are determined in the fit of dt distribution:*

$$\wp_{sig}(\Delta t') = \frac{exp\left(-\frac{|\Delta t'|}{\tau}\right)}{4\tau}\,[1 + q(A\,cos(\Delta m\,\Delta t') + S\,sin(\Delta m\,\Delta t')\,)]$$

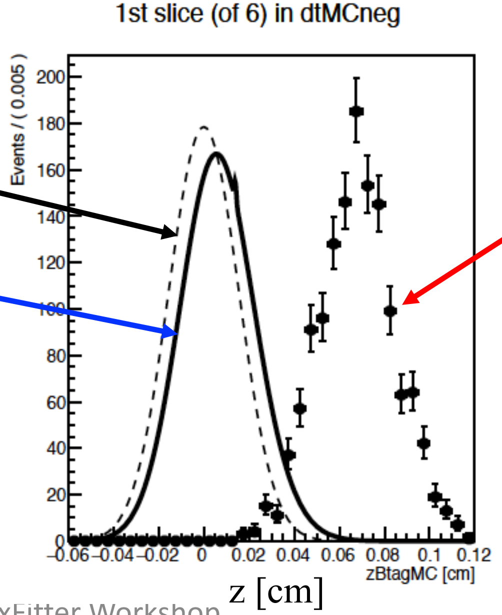
Beam spot and spatial distributions of Btag decays

→ z in lab is considered for simplicity, the boost direction would be more appropriate



Y(4s) beam spot

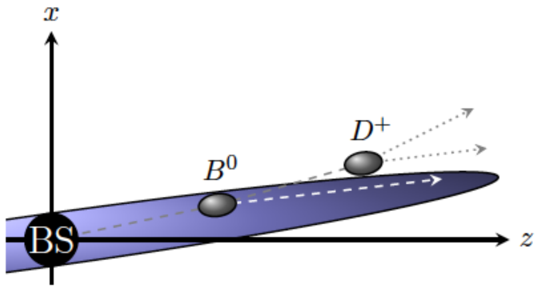
Btag is first
$\Delta t' > 0$



example of large decay length
Btag is second
-8 ps $< \Delta t' < -6.7$ ps

→ very far away in z !

Tiny size of the beam spot in Belle II (comparable with B0 lifetime) is an advantage but also a challenge: too intensive use of this information in reconstruction leads to dependence of the reco results on spatial position of the „second" Btag and on physics parameters

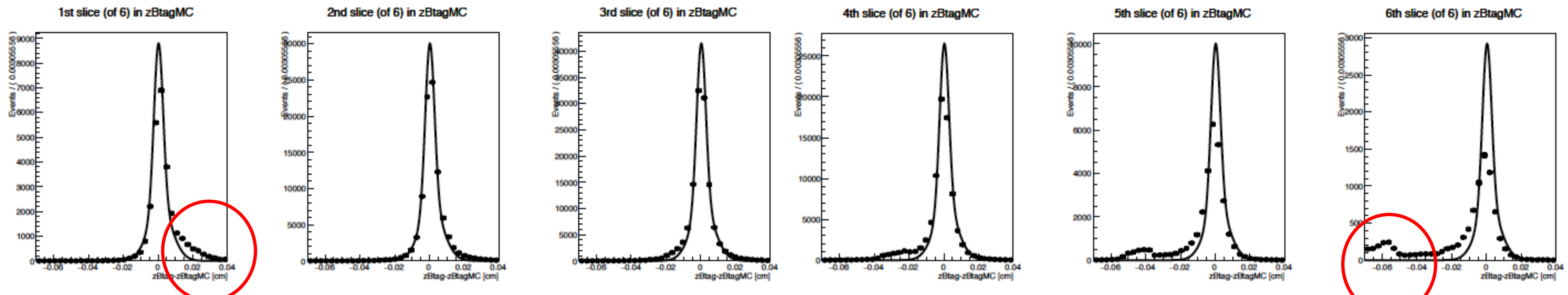


Resolution in zBtag and in dt as function of zBtag(true) and dt(true)

6 slices in true z(Btag) from -300 μm to 700 μm

- solid line is a three gauss fit to all MC events

Δ(zBtag)
(reco – true)



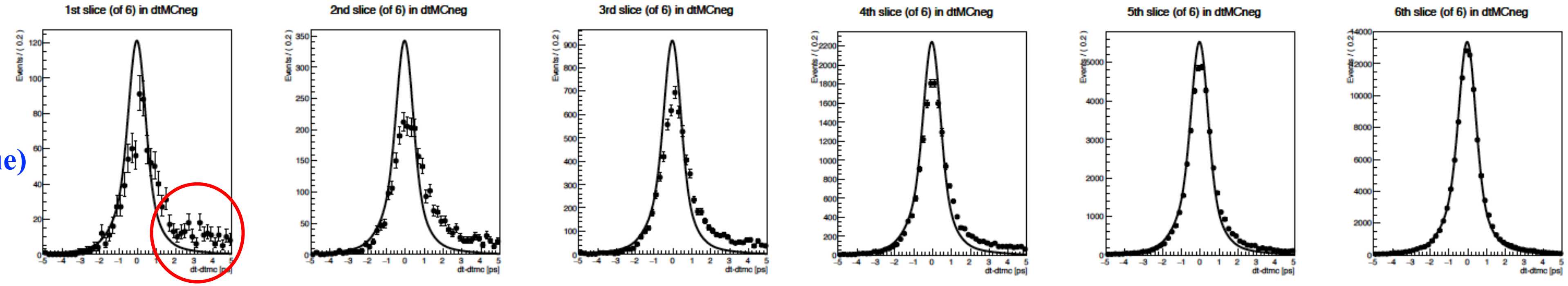
Δz [cm]

→ strong dependence of z-reconstruction on the true absolute z position of Btag

6 slices for negative true $\Delta t'$ from -8 ps to 0

- solid line is a three gauss fit to MC events with positive dt

Δ(dt)
(reco – true)



$(\Delta t\text{-}\Delta t')$ [ps]

→ strong dependence of the Δt resolution on true $\Delta t'$ for $\Delta t'$<0 !!!

The MPI concept with pdfs determined on fly from MC samples

Prerequisite: *should cope with dependencies of the resolution function on physics parameters to be measured*
should use full information for every event in a data sample → e.g. an unbinned maximum likelihood method

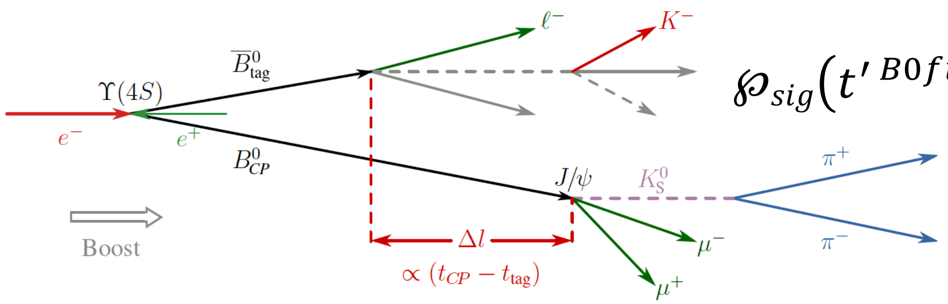
Generation+Simulation+Reconstruction *for production of MC samples is the best and the only direct way in hand to make convolution of the underlying physics distributions and the detector effects (resolution functions). The "dt resolution functions" used so far are approximations of detector effects derived from MC samples.*
→ use the Generation+Simulation+Reconstruction convolution in full glory.

The MPI concept is based on re-weighting of mc events:

- *probability of every event in data is calculated on fly using weighted sample of MC events*
- *MC events, re-weighed by the ratio Pdt(new par.)/Pdt(gen par.), are equivalent to new simulation with new parameters*
- *event probabilities are smooth and continuous functions of physics parameters defined everywhere: well suited for fitting*

- *in the original MC method widely used at HERA and LHC, expected number of MC events are directly compared with number of data events and, therefore, very good description of detector effects in simulation of MC is needed*
- *in the MPI approach it is not obligatory:*
  external parameters can be used for evaluation of quality of dt reconstruction which distributions could differ for data and MC, for example uncertainty on dt or external parameters similar to "ξ"("h") used in Belle

Central idea: re-weighting of MC events instead of new simulations



$$\wp_{sig}\left(t'^{\,B0first}, \Delta t'\right) = \frac{exp\left(-\frac{\left|t'^{\,B0first}\right|}{\tau}\right)}{4\tau}\,\frac{exp\left(-\frac{\left|\Delta t'\right|}{\tau}\right)}{4\tau}\left[1 + q\left(A\,cos(\Delta m\,\Delta t') + S\,sin(\Delta m\,\Delta t')\right)\right]$$

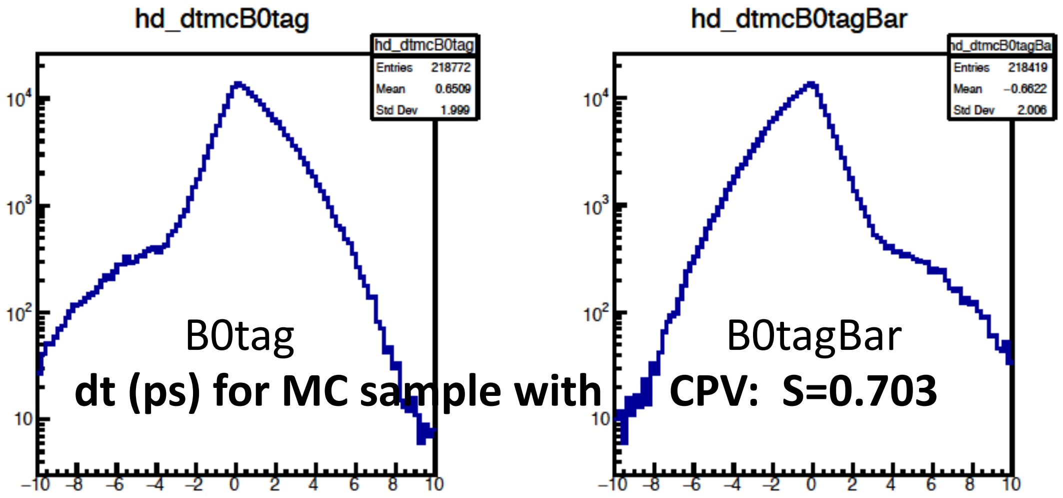
- *only two generated quantities serve as input to simulation:*
  spatial positions of the first and the second B0s from Y(4S)→B0B0bar
- *these two positions define the whole kinematics of the event*

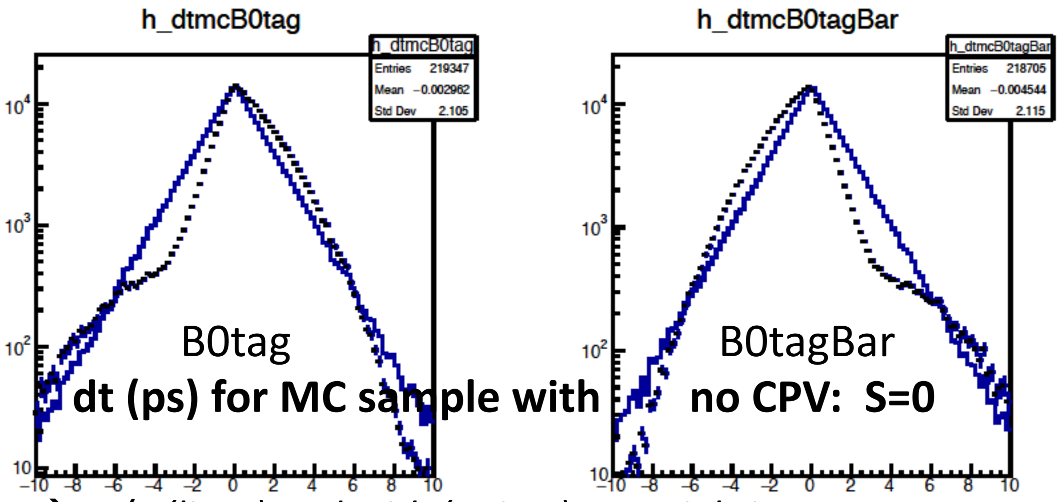
- *these positions are determined by the underlying physics quantities:* $t^{B0fisrt}$ *for decay of the first B0 and dt=t_B0sig–t_B0tag*
- *only these two physics quantities depend on the theory parameters tau(B0), dm, A and S , q=+1(B0tag),q=-1(B0tagBar)*

→ *decisive is to have correct distributions on these physics quantities for new set of parameters of interest*

it can be achieved in ***two (equivalent) ways****: either* ***by production of a new simulation*** *or*

by weighting of old simulation with ratio of Psig for new and old physics parameters as function of generated $t^{B0fisrt}$ ***and dt***





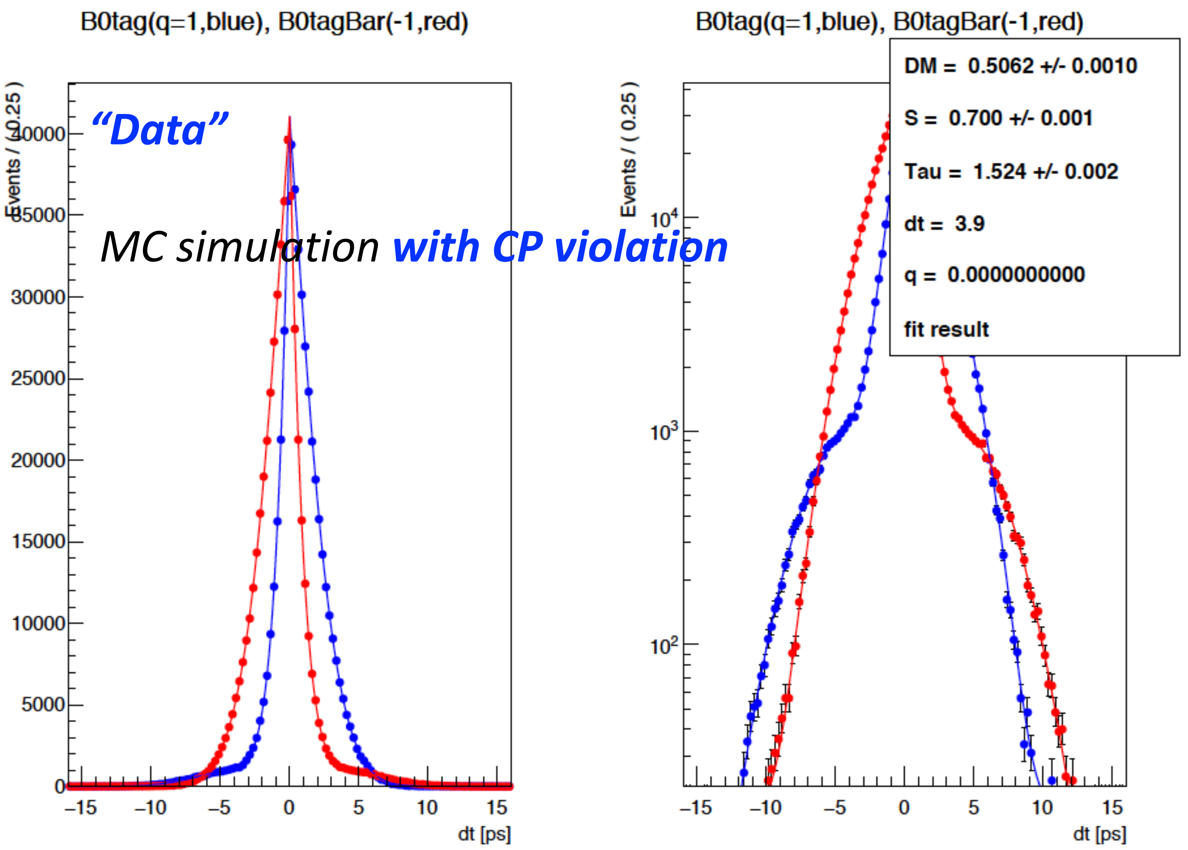
→ *w/o (lines) and with (points) re-weighting to S=0.703*

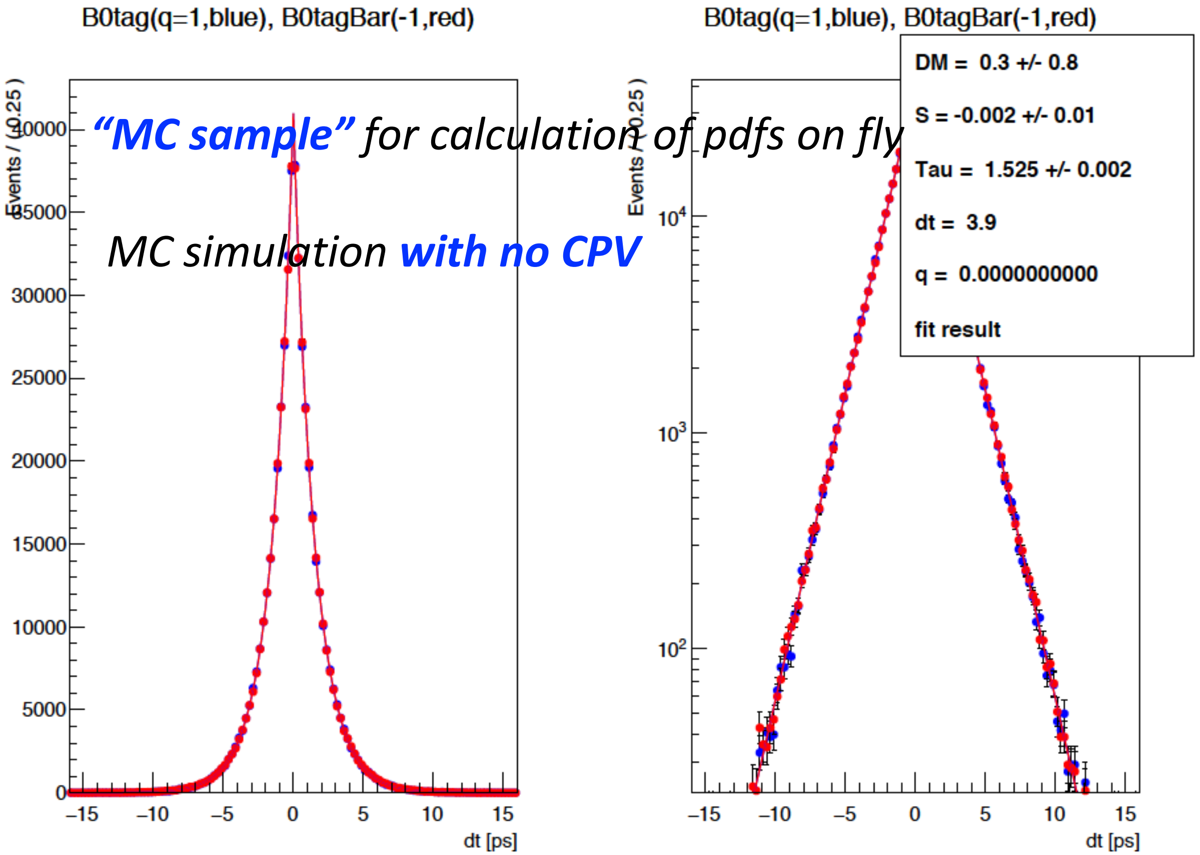
Nice features of the approach

- *variation of input physics parameters by weighting → no need for templates or new simulations !*
- *result is obtained in one go → no need for iteration procedure which would be required otherwise*
- *(known&possible) correlations with physics parameters are automatically taken into account*
- *no problem with analytical description of the dt resolution shapes, they are taken directly from simulation.*
- *biases related to tracks from charm and K0 included in Btag vertex fitting are taken into account*
- *effects due to approximations used in calculation of reconstructed dt are taken into account*
- *straightforward treatment of efficiencies*
- *easy change of selection conditions with immediate modification of the event probability functions*
- *possibility to keep all advantages of the old method making use of external parameters on event by event basis by fitting dt shapes in slices of external parameters with their distributions taken from data (conditional pdfs)*
- *possibility to apply on fly "adjustments" of the MC response (e.g. additional smearing) and efficiencies, may be even with an additional free parameter which could allow better description of data by simulation*
- *universality: applicable to any process, if proper selection criteria and MC samples are defined*
- *weighting technique is much cheaper than convolutions of complicated functions*

Two MC files serving as “Data” and “MC sample for reweighting”

***Two MC samples with and without CPV**:* $\Upsilon(4S) \rightarrow B^0 B^0,\ B^0 \rightarrow J/\psi(\mu\mu)\ K_S^0$ *(1 mln events each sample)*





MC with CPV: *fit of generated dt=tB0sig−tB0tag*

τ = 1.524 ±0.002 ps, Δm = 0.5062 ± 0.0010 ps^{-1}, S = 0.700 ± 0.001

→ expected from steering:

τ = 1.525 ps Δm = 0.507 ps^{-1} S = 0.703 (2beta=0.78)

MC with no CPV: *fit of generated dt=tB0sig−tB0tag*

τ = 1.524 ±0.002 ps, *Δm = 0.3 ± 0.8 ps^{-1}, S = 0 ± 0.01*

→ expected from steering:

τ = 1.525 ps *Δm = 0 S = 0*

Technical details of the MPI approach

Fitting platform: *ALPOS (https://github.com/zleba/alpos)* *[RooFit does not provide needed functionality]*

- Daniel Britzger (MPI) helped to set up the fit in ALPOS, which is actively used for various fits (jets, …) at HERA and LHC

***The fit**: (unbinned) maximum likelihood method*

***Two samples**: "Data" (simulation with CP violation) and "MC sample for pdf calculations on fly" (simulation with no CPV)*
***Loose selection** (same for both samples) with MC matching and requirement of at least one PXD hit for muons from J/ψ*

***Grid in reconstructed dt** (104 bins): -100 -38 -22 -14 -10 -8 -7 -6 -5.5 -5 -4.5 -4 -3.75 -3.5 -3.25 -3 -2.8 -2.6 -2.4 -2.2 -2 -1.9 -1.8 -1.7 -1.6 -1.5 -1.4 -1.3 -1.2 -1.1 -1 -0.9 -0.85 -0.8 -0.75 -0.7 -0.65 -0.6 -0.55 -0.5 -0.45 -0.4 -0.35 -0.3 -0.25 -0.2 -0.15 -0.1 -0.05 0 + similar in positive direction*
***Grid in ddt** (uncertainty of dt), 10 bins: ddt_binning: 0 0.3 0.4 0.5 0.6 0.7 0.8 1.0 1.3 2.0 10*
***Flavor tagger output**: qrfltag (B0_FBDT_qrCombined) → defines 7 regions and q=sign(qrfltag)=1(B0fltag),-1(B0fltagBar)*
7 bins : abs(qrfltag)-binning: 0 0.1 0.25 0.50 0.625 0.75 0.875 1

***14 classes in total**: 7 regions in abs(qrfltag) and. q=1,-1 (for data) or qmc = 1 (B0tag), -1 (B0tagBar) (for MC)*
→ could (will) be extended for new external parameters(classes) e.g. "with/wo PXD hits", "ξ" or "h" similar to Belle, …

Probability calculation for given event in “data”

initial step :

1. *loop over samples of data and MC events and fill the following information*

   - “data”: fill 2D array with numbers of selected data events in bins (idt,iddt) for each qrfltag in ic-bins (according to grids)

   *- “MC sample”: keep in memory 5 variables for selected MC events - idt, iddt, icmc=qmc*abs(ic), tmcB0first, dtmc*

   here, ic=-7,..,7; qmc=-1 (B0tag),+1(B0tagBar)

at each iteration in minuit :

1. *loop over MC events in memory and weight each MC event by the ratio of Psig for Tau,dm,A,S of the current iteration and Tau0=1.525 ps & S=A=0 used in the generation of the MC sample with “no CPV”*
2. *fill 2D array (idt, iddt) with sums of weights of MC events for each class icmc (= integration over tmcB0first & dtmc)*
3. *calculation of probability for each given data event with (idt,iddt,ic), calculated independently in bins (iddt,ic)*

$$\text{Prob\_ev(idt)} = \{\, [\text{n\_pos(idt)}+\text{n\_neg(idt)}] + \text{sign(ic)}*(1-2W_{\text{abs(ic)}})[\text{n\_pos(idt)}-\text{n\_neg(idt)}] \,\} / \{\, (\text{sum n\_pos}) + (\text{sum n\_neg}) \,\}$$

   n_pos, n_neg are for positive and negative icmc with abs(icmc)=abs(ic); sum n_pos (sum n_neg) are sums over idt

4. *$W_{1,..,7}$ - seven delusion factors (probabilities to make wrong decision) for 7 regions in abs(qrfltag), i.e. for abs(ic)-bins*
5. *supply minuit with sum {-2 log Prob_ev(idt,iddt,ic) } - sum over all events in the “data” sample*

→ *10 free parameters: Tau, dm, S, (A=0), $W_{1,...,7}$*

- sometimes to ensure 0<W<1, Wpar are fitted instead of W → W = 0.5{ - sign(Wpar)[1-exp(-0.5-abs(Wpar))/exp(-0.5)]}*

- sometimes to ensure -1<S<1, Spar are fitted instead of S → S = sign(Spar)[1-exp(-0.5-abs(Spar))/exp(-0.5)]*

- it is equivalent to MC method if the fit of dt shapes is performed for different classes with free normalization of each class

- classes “ic” could be easily extended for further external parameters

Prove of the MPI concept

Fit converges quickly: < 60 sec with room for further optimization of the code and parallelization of calculations
Data: 437191 good event (out of 1027869 ev); MC sample: 438052 good events (out of 1028551 ev)
10 free parameters, initial values are far from expectations: Tau=1.4, dm=0.4, S=0.5, W(1-7)=0.3

MinFCN = 3.84699e+06 NDf = 0 Edm = 6.15154e-06 NCalls = 74 Runtime 00:00:10 ***only one class, perfect "flavor tagging": sign(qrfltag)=qmc***

BelleTDCPV.S	=	*0.699429 +/- 0.00265984*	***0.699+-0.003***	***0.703 expected***	***→ perfect agreement (within one sigma)***
BelleTDCPV.tau	=	*1.52403 +/- 0.00292701*	***1.524+-0.003***	***1.525 expected***	***→ perfect performance (10 sec)***
BelleTDCPV.dm	=	*0.506841 +/- 0.00211855*	***0.507+-0.002***	***0.507 expected***	***→ prove of the MPI concept***

Conclusion: ***The MPI concept works perfectly !***

Work in still progress:
make use of the flavor tagger, backgrounds to be included, …

Realistic exercise (although without backgrounds) with
7 classes and qrfltag from flavor tagger (10 free parameters)
→ reasonable results and very fast: Runtime 00:01:03 (63 sec !)

Conclusions

- The MPI concept of a time-dependent fit is adequate to new challenges in Belle II related to improved precision of PXD and the size of the beam spot.

- It is demonstrated that the method based on re-weighting of MC events and calculation of probabilities for every data event "on fly" works well for realistic conditions (accurate, fast and efficient). The work is still in progress: flavor tagger, backgrounds are to be included , …

Input and expertise are welcomed:

- advantages/disadvantages of unbinned maximum likelihood methods w.r.t. minimisation of chi2

- uncertainties due to MC statistics (statistics in tails is low for any MC sample)

- how to treat limits ($-1<S<1$, $0<W<1$), if function – which function is better ?

- are there complications with efficiencies due to loss of MC events which can not be therefore re-weighted?

- further optimizations of the code and parallelization of calculations

- easy way to include many pdfs which are on the market: Breit-Wigner, Argus, …

- the fitting task is simple, straightforward and stand alone:

- how to make the fitting package portable to "any calculator" with running Root ?

- make it (as much as possible) user friendly and convincing for users in Belle II

- output services: histos, comparisons, …