Flavour Tagging Tutorial

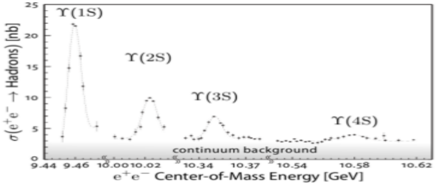
Stefano Lacaprara[1], Fernando Abudinén[2], Luigi Ligioi[2]

stefano.lacaprara@pd.infn.it

[1]INFN Padova, [2]Max-Plank-Institute für Physik

Belle II Starter Kit Workshop,
KEK, 11 October 2018

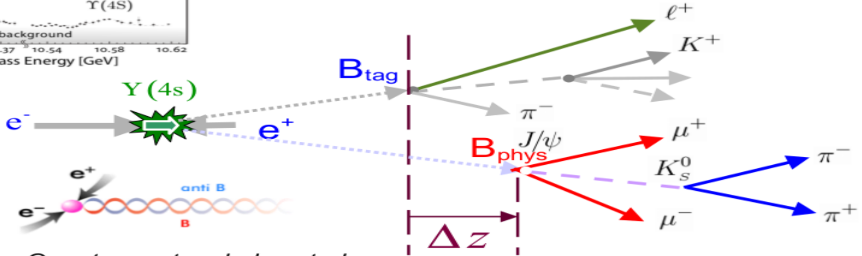
Time Dependent CP violation in a nutshell



- Y(4S) is the first resonance just above the $B\overline{B}$ production threshold
- Only $B\overline{B}$ pairs are produced, and are at rest in the Y(4S) frame

$$\Delta t = \frac{\Delta z}{\beta \gamma c}$$

Resolution on Δt will be dominated by the resolution of the tagging side vertex

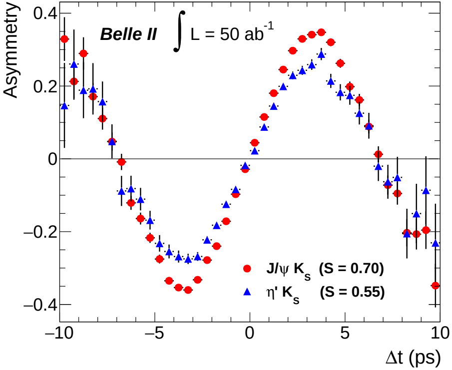


Quantum entangled neutral B meson pair production

Belle ~ 200 μm
Belle II ~ 130 μm

Δt probability parametrization $\mathcal{P}(\Delta t, q) = \frac{e^{-|\Delta t|/\tau_{B^0}}}{4\tau_{B^0}}\left[1 + q\left(\mathcal{A}_{CP}\cos\Delta m_d \Delta t + \mathcal{S}_{CP}\sin\Delta m_d \Delta t\right)\right]$

Motivations



Needs everything!

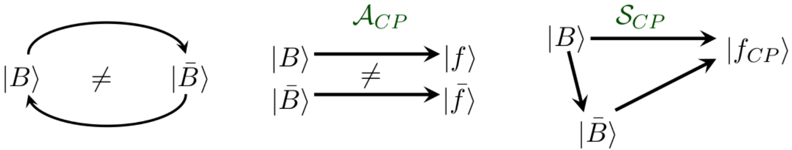
- exclusive B^0 signal reconstruction;
  - charged and neutral particles
  - PID
  - vertexing
- Flavour tagging of B^0_{tag};
- Measure Δz

Physics motivation

- determination of $\phi_1(\beta)$
- measurement of $\phi_2(\alpha)$
- sensitiveness to New Physics

$$Asym = \frac{\mathcal{P}(B^0 \to X_{CP}) - \mathcal{P}(\bar{B}^0 \to X_{CP})}{\mathcal{P}(B^0) + \mathcal{P}(\bar{B}^0)} = \mathcal{A}\cos(\Delta m \Delta t) + \mathcal{S}\sin(\Delta m \Delta t)$$

Some definition



- Direct $\mathcal{A}_{CP}^{J/\psi K_S^0} = 0$, Mix-induced $\mathcal{S}_{CP}^{J/\psi K_S^0} = \sin(2\phi_1)$
- FT is possible only for a fraction of events: efficiency ϵ (very high, $\epsilon \sim 99\%$)
- a fraction w of them is wrongly classified

$$\mathcal{P}^{Obs}(\Delta t, q, \epsilon, w) = \epsilon \left[(1-w)\mathcal{P}^{Sig}(\Delta t, q) + w\mathcal{P}^{Sig}(\Delta t, -q)\right]$$

$$= \frac{e^{-|\Delta t|/\tau}}{4\tau}\epsilon\left\{1 + q(1-2w)\cdot\left[\eta_{CP} S_{CP}\sin(\Delta m \Delta t) + A_{CP}\cos(\Delta m \Delta t)\right]\right\}$$

- q is flavour, $\eta_{CP} = \pm 1$ is CP final state; τ, Δm from PDG
- $r = (1-2w)$ is called dilution factor

Statistical significance

$$N^{tag}_{B^0} = \varepsilon(1-w)N_{B^0} + \varepsilon w N_{\overline{B}^0}$$

$$N^{tag}_{\overline{B}^0} = \varepsilon(1-w)N_{\overline{B}^0} + \varepsilon w N_{B^0}$$

$$Asym^{Obs}(\Delta t) = \frac{N^{tag}_{B^0} - N^{tag}_{\overline{B}^0}}{N^{tag}_{B^0} + N^{tag}_{\overline{B}^0}} = (1-2w)\cdot\frac{N_{B^0} - N_{\overline{B}^0}}{N_{B^0} + N_{\overline{B}^0}} = r\cdot Asym^0$$

- Stat uncert on $Asym^0$ is $\sigma_{Asym^0} = \frac{\sigma_{Asym^{obs}}}{r}$;
- If $N_{tag} = N^{tag}_{B^0} + N^{tag}_{\overline{B}^0}$, with small asymmetry ($N^{tag}_{B^0} \approx N^{tag}_{\overline{B}^0}$): $\sigma_{Asym^{obs}} \propto \frac{1}{\sqrt{N_{tag}}} \propto \frac{1}{\sqrt{\varepsilon}}$
- $\sigma_{Asym^0} = \sigma\mathcal{S}_{CP} = \sin 2\phi_1 \propto \frac{1}{r\sqrt{\epsilon}} = \frac{1}{\varepsilon_{eff}}$.
- effective tagging efficiency $\varepsilon_{eff} = \varepsilon r^2$

the statistical uncertainty on $Asym^0$ of a sample of N_B candidates, with tagging eff ε and wrong tag probability w is equivalent to that of a sample of $\varepsilon_{eff} N_B$ with perfect tagging.

It is more complicated than this ...

- The wrong tag fraction is typically different for B^0 and $\overline{B}^0$ Δw
- We typically use many tagging categories (13 today in Belle II)
  - $\varepsilon_{eff} = \Sigma_i \varepsilon_i \langle r_i \rangle^2$
- the best sensitivity on $\mathcal{S}_{CP}$ is obtained form a UML fit to data

$$\mathcal{P}^i_j = \underbrace{\mathcal{T}_j\left(\Delta t^i, \sigma^i_{\Delta t}, \eta^i_{CP}\right)}_{\text{time-dep part}} \prod_k \underbrace{\mathcal{Q}_{k,j}(x^i_k)}_{\text{time integrated}}$$

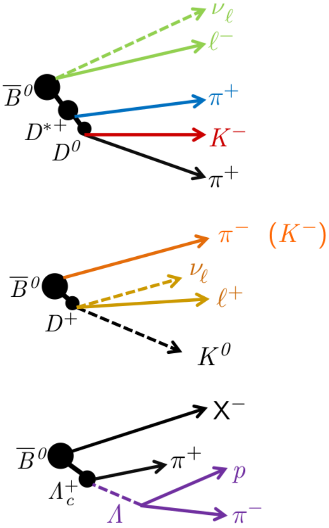
- $\varepsilon_{eff} = \Sigma_i \varepsilon_i {r_i}^2 \left(1 + \frac{12 x_i^2 r_i^2 \mathcal{S}^2_{CP}}{1 + 16 x_i^2}\right)$, with $x_i = \Delta m / \Gamma_i$
- as before for $\mathcal{S}_{CP} = 0$.
- For large $\mathcal{S}_{CP} \approx 1$ can be as large as 60% [Cahn[2000],Le Diberder[1990]]

How do we Flavour Tag an event?

- Full reconstruction of signal side. Eg. $B^0 \to J/\psi K^0_S$ or $B^0 \to \phi K^0_S$ or …;
- Perform full reconstruction of tag B (Rest Of Event: all that is not signal side);
- Large fraction of B decays is flavour specific, namely can only be reached through a decay of b quark or via a $\overline{b}$ quark.
  - eg $B^0 \to D^{*-}\ell^+\nu_\ell$ the charge of ℓ identify the flavour of B^0, as long as ℓ is coming from B^0 and not from secondary D decay
  - So we need to identify the ℓ but that is not enough.
  - Need to look at kinematic variables to understand if from B^0 or from D (using MVA technique)
- But so many B^0 decays are possible that also inclusive technique are also used
- In BelleII we use 13 different categories, both inclusive and exclusive.
- Very important!: **Flavour Tagger** is candidate-based (not event-based!)
  - If you have multiple B^{Sig} reconstructed in an event (eg decays into neutrals $B^0 \to \eta'(\to \eta(\to \gamma\gamma)\pi^+\pi^-)K^0_S(\to \pi^0\pi^0))$
  - you have a different ROE for each candidate
  - and a different flavour tag for each candidate

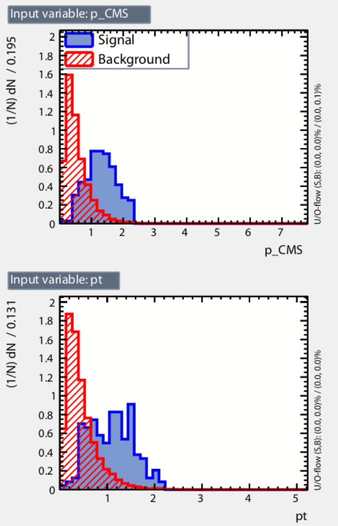
Flavour Tagger categories

Categories	Targets
Electron	e^-
Intermediate Electron	e^+
Muon	μ^-
Intermediate Muon	μ^+
KinLepton	e^-
Intermediate KinLepton	ℓ^+
Kaon	K^-
KaonPion	K^-, π^+
SlowPion	π^+
FastHadron	π^-, K^-
MaximumP	ℓ^-, π^-
FSC	ℓ^-, π^+
Lambda	Λ
Total= 13	



Leptonic Categories

- Can be primary ($b \to c\ell^{-}\nu$)
- or secondary ($b \to c \to s\ell^{+}\nu$) leptons, with opposite charge
- Separate primary from cascade leptons using p and p_T spectrum (harder for primary)
- other variables:
  - E_W^{90}: energy on the hemisphere defined by direction of virtual W boson.
  - p^*_{miss} and $\cos\theta^*_{miss}$
  - $\cos\theta^*_T$ angle between ℓ and thrust axis of $B^0{}_{tag}$
  - M_{recoil} inv mass of the recoil system
- considerate separately `Electron` and `Muon`
- if identified as cascade, `IntermediateElectron` and `IntermediateMuon`
- inclusively in `KinLepton` and `IntermediateKinLepton`
  - of course inclusive and exclusive categories are strongly correlated.
  - still inclusive categories improve overall performances



Hadronic Categories

- **Kaon**
  - Dominant decay $B^0 \to D(\to K^- X)X$ tag from K charge.
  - K multiplicity 0.78 ± 0.08[PDG]
  - right(wrong) sign K is 0.58(0.13)[Albrecht et al[1994b]]
  - very powerful source of tagging info!
    - ★ considered also $n_{K_S^0}$ in ROE
    - ★ K_S^0 from $b \to c\bar{c}s$ decay or $s\ \bar{s}$ from gluon splitting
    - ★ other kin variables as for leptons
  - combine the three K with highest $q \cdot r$
- **SlowPion**
  - $\pi^\pm$ from $D^{*\pm}$ decays.
  - same variables as `Kaon`
  - $\cos\theta_T^*$ particularly powerful
- **KaonPion**
  - Correlation between slow pion and kaon
- **MaximumPstar**
  - very inclusive tag by looking at charge of the highest CMS-momentum particle in the ROE
  - such as from hadronization of W or leptons
  - high $\varepsilon \approx 100\%$: fails only if no tracks in ROE
  - ε_{eff} not as good.
- **FSC**
  - Inclusive tagger using correlation between fast (MaximumPstar) and slow (SlowPion) particles
- **Lambda**
  - From $b \to c \to s$ decay, $\Lambda \to p\pi$
  - clean tagger but with low ε

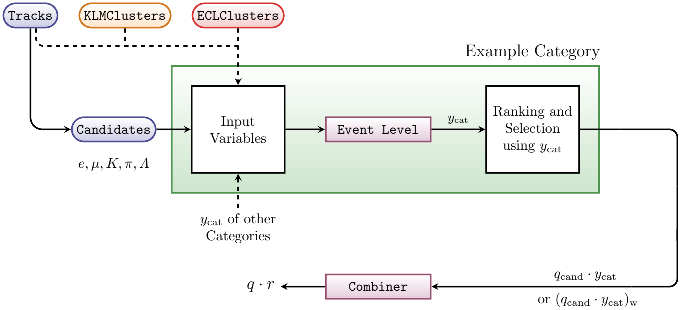
Categories	Discriminating input variables
Electron Int. Electron	$\mathcal{L}_e$, p^*, p_t^*, p, p_t, $\cos\theta$, d_0, $\lvert x\rvert$, M_{rec}^2, E_{90}^W, p_{miss}^*, $\cos\theta_{miss}^*$, $\cos\theta_T^*$, p-val.
Muon Int. Muon	$\mathcal{L}_\mu$, p^*, p_t^*, p, p_t, $\cos\theta$, d_0, $\lvert x\rvert$, M_{rec}^2, E_{90}^W, p_{miss}^*, $\cos\theta_{miss}^*$, $\cos\theta_T^*$, p-val.
Kin. Lepton Int. Kin. Lep.	$\mathcal{L}_e$, $\mathcal{L}_\mu$, p^*, p_t^*, p, p_t, $\cos\theta$, d_0, $\lvert x\rvert$, M_{rec}^2, E_{90}^W, p_{miss}^*, $\cos\theta_{miss}^*$, $\cos\theta_T^*$, p-val.
Kaon	$\mathcal{L}_K$, p^*, p_t^*, p, p_t, $\cos\theta$, d_0, $\lvert x\rvert$, $n_{K_S^0}$, $\sum p_t^2$, M_{rec}^2, E_{90}^W, p_{miss}^*, $\cos\theta_{miss}^*$, $\cos\theta_T^*$, χ^2
Slow Pion Fast Hadron	$\mathcal{L}_\pi$, $\mathcal{L}_{e^-}$, $\mathcal{L}_K$, p^*, p_t^*, p, p_t, $\cos\theta$, d_0, $\lvert x\rvert$, $n_{K_S^0}$, $\sum p_t^2$, M_{rec}^2, E_{90}^W, p_{miss}^*, $\cos\theta_{miss}^*$, $\cos\theta_T^*$, p-val.
Kaon-Pion	$\mathcal{L}_K$, y_{Kaon}, $y_{SlowPion}$, $\cos\theta_{K\pi}^*$, $q_P K \cdot q_\pi$
Maximum P*	p^*, p_t^*, p, p_t, d_0, $\lvert x\rvert$, $\cos\theta_T^*$
FSC	$\mathcal{L}_{KSlow}$, p_{Slow}^*, p_{Fast}^*, $\cos\theta_{T,\,Slow}^*$, $\cos\theta_{T,\,Fast}^*$, $\cos\theta_{SlowFast}^*$, $q_{Slow}\cdot q_{Fast}$
Lambda	$\mathcal{L}_p$, $\mathcal{L}_\pi$, p_Λ^*, p_Λ, p_p^*, p_p, p_π^*, p_π, q_Λ, M_Λ, $n_{K_S^0}$, $\cos\theta_{x_\Lambda,p_\Lambda}$, $\lvert x_\Lambda\rvert$, σ_Λ^{zz}, p-val.

Five particle lists: e^-, μ^-, K, π, p (Λ from p and π).

200 variables, each computed once for each particle: exclusive 108 variables

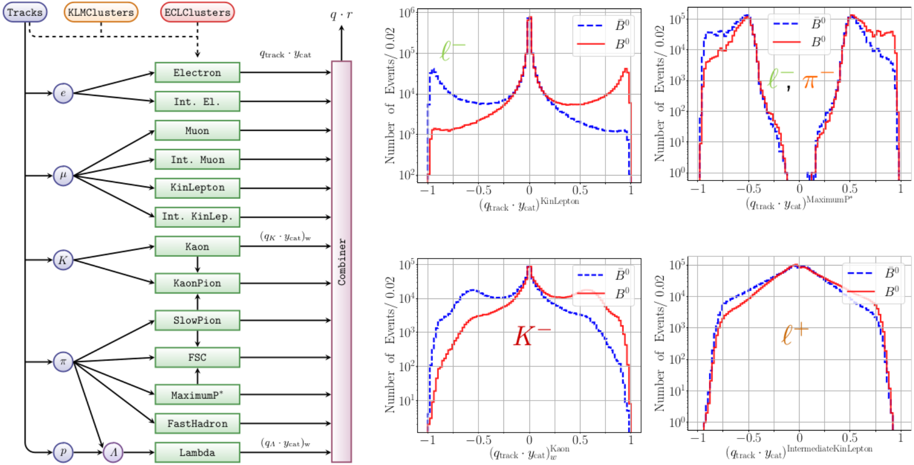
Category tagger: procedure

- Starting info: Objects in the Tag Side (ROE)
- possibly also from other categories (eg KaonPion)
- MVA (FBDT) to determine the tag flavour and tag probability $y_{cat} = q \cdot r \in [-1, 1]$
- might have different target for each category (eg different K or ℓ)
  - consider that with higher y_{cat}
- Training of MVA on $B^0 \to J/\psi K^0_S$

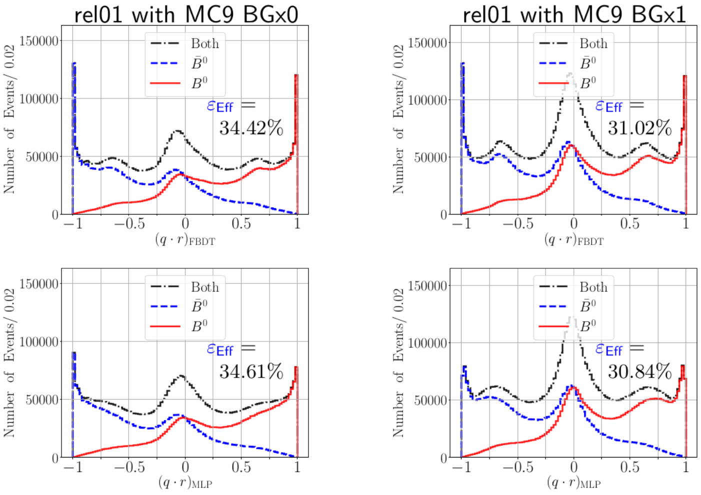


y_{cat} for each category is then passed to the **Combiner** which return a single value of $q \cdot r$ for each candidate.

Combining all together

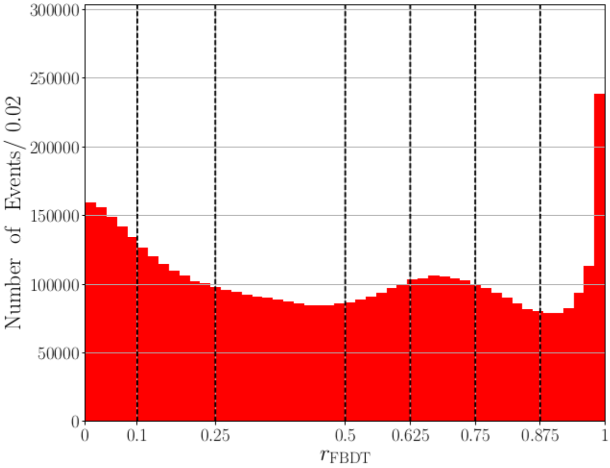


Performances with MC9 (release-01-xx-xx)



FBDT used as default, FANN (MLP) as a cross check

Effective efficiency ε_{eff} calculation

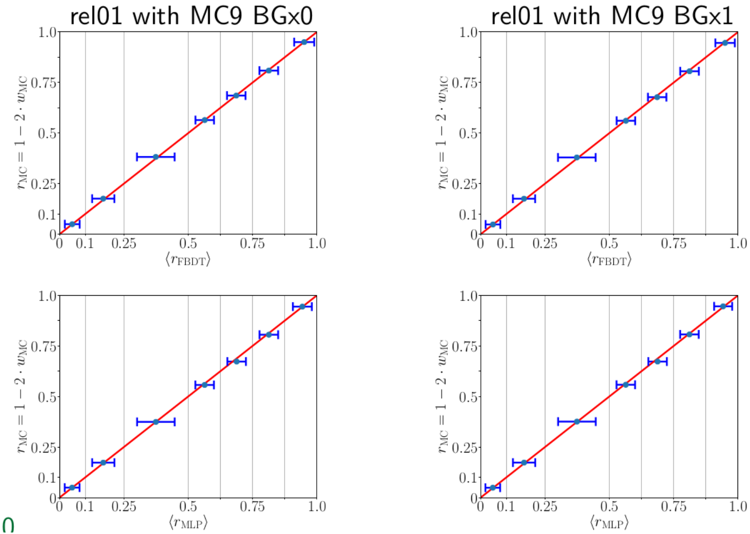


- Binning ⇒ correction with real data!
- Efficiency:

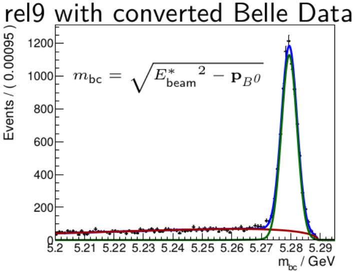
$$\varepsilon_{\text{Eff}} = \sum_i \varepsilon_i \cdot \langle r_i \rangle^2$$

- $r_{\text{MC}} = 1 - 2 \cdot w_{\text{MC}}$
- Calibration: r_{MC} linear to r_{Output}

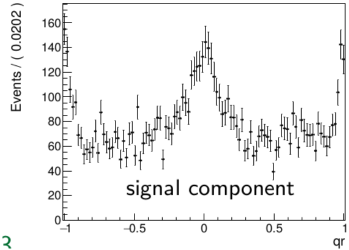
Linearity wrt MC truth

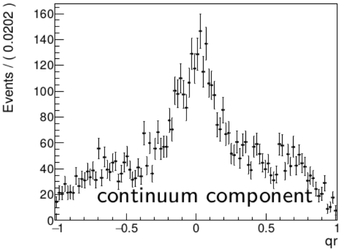


10

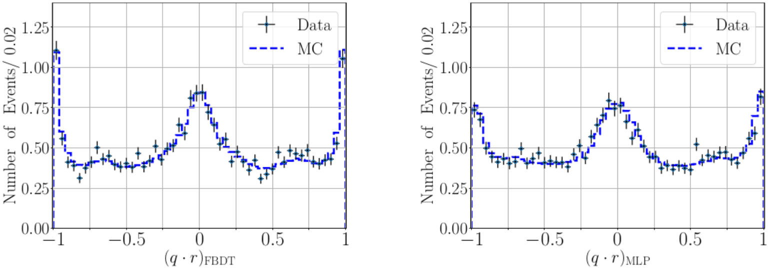


- Splot performed with converted Belle data using m_{bc} as discriminating variable.
- Full Belle 0.8 ab^{-1} $B^0 \rightarrow J/\psi K^0_S$





- Belle Data distribution weighted with splot output variable (signal component).



- Nice overlap of converted Belle MC and data ☺.
- $\varepsilon_{\mathsf{Eff}} = 34.2\%$ on converted Belle MC (Belle $\sim 29\%$) ☺.
- Deep neural network flavor tagger reached 34.4% on converted Belle MC.

Example and documentation

- Belle2 Starter-Kit
  - Confluence
  - stash
  - We will go through B2T_Advanced_3_FlavorTagger now
- Tutorial script
  - `analysis/examples/tutorials/B2A801-FlavorTagger.py`
- Tutorial for use with B2BII converted MC or data:
  - `analysis/examples/tutorials/B2A801-FlavorTagger-BelleMC.py`
- Training and validation plots performed at `kekcc` with the script `flavorTaggerValidatorInParalell.sh` in `/home/belle2/abudinen/public/release1ValidationScripts/`
  - Example if you want to train by yourself
- Confluence page:
  - https://confluence.desy.de/display/BI/Physics+FlavorTagger
- B2TIP report

Additional stuff

Additional or backup slides

Diberder, F. L. (1990). Precision on CP -Violation Measurements and Requirement on the Vertex Resolution. Technical Report BaBar Analysis Document #34.

H. Albrecht et al (1994). Kaons in flavor tagged B decays. Z.Phys *C62*.

R.Cahn (2000). TagMixZ and its Application to the Analysis of CP Violation. Technical Report BaBar Analysis document #17.