# System8 for
# B tagging algorithms calibration

Thomas Delemontex (LPSC)

LPSC-CPPM / ATLAS group meeting 06/09/2010

# Plan

- System used at LPSC
- Kappas determination
- Error propagation in System8
- Some further ideas and comments

# The system used in Grenoble

$$
\left\{
\begin{array}{rcrcl}
f_b & + & f_{cl} & = & 1 \\
\varepsilon_b^{(X)} f_b & + & \varepsilon_{cl}^{(X)} f_{cl} & = & q^{(X)} \\
\varepsilon_b^{(Y)} f_b & + & \varepsilon_{cl}^{(Y)} f_{cl} & = & q^{(Y)} \\
\varepsilon_b^{(Z)} f_b & + & \varepsilon_{cl}^{(Z)} f_{cl} & = & q^{(Z)} \\
\kappa_b^{(X,Y)} \varepsilon_b^{(X)} \varepsilon_b^{(Y)} f_b & + & \kappa_{cl}^{(X,Y)} \varepsilon_{cl}^{(X)} \varepsilon_{cl}^{(Y)} f_{cl} & = & q^{(X,Y)} \\
\kappa_b^{(Y,Z)} \varepsilon_b^{(Y)} \varepsilon_b^{(Z)} f_b & + & \kappa_{cl}^{(Y,Z)} \varepsilon_{cl}^{(Y)} \varepsilon_{cl}^{(Z)} f_{cl} & = & q^{(Y,Z)} \\
\kappa_b^{(Z,X)} \varepsilon_b^{(Z)} \varepsilon_b^{(X)} f_b & + & \kappa_{cl}^{(Z,X)} \varepsilon_{cl}^{(Z)} \varepsilon_{cl}^{(X)} f_{cl} & = & q^{(Z,X)} \\
\kappa_b^{(X,Y,Z)} \varepsilon_b^{(X)} \varepsilon_b^{(Y)} \varepsilon_b^{(Z)} f_b & + & \kappa_{cl}^{(X,Y,Z)} \varepsilon_{cl}^{(X)} \varepsilon_{cl}^{(Y)} \varepsilon_{cl}^{(Z)} f_{cl} & = & q^{(X,Y,Z)}
\end{array}
\right.
$$

Exactly the same system as the "n-p" system excepted for :

→ Equations normalized by the number of selected jets (without tagger applied) N
so: $Q^{(\dots)} = n^{(\dots)} / N$ and $f_b = n_b / N$

→ Equations with p become equations with $\varepsilon^Z$

→ Recreate the system symmetry between all taggers (opposite tagged jet is a tagger)

→ Maybe a difference in the $K^{123} \neq \alpha_{7\text{-}8}$ definition

# Kappas

**Kappa determination is crucial → Can introduce great discrepancies in final results**

**Kappas are determined using MC Samples**

In our solver 8 kappas (correlation factors) are considered :

$K^{12}$ , $K^{23}$ , $K^{31}$ , $K^{123}$ for signal and backgrounds so correlation between tagger1 and tagger2 , …

$K^{12} = \varepsilon^{12}/(\varepsilon^{1} * \varepsilon^{2})$

$K^{123} = \varepsilon^{123}/(\varepsilon^{1} * \varepsilon^{2} * \varepsilon^{3})$ ← $\alpha_7 = \varepsilon^{123}/ (\varepsilon^{23} * \varepsilon^{13}) = \varepsilon^{123}/ (\alpha_6 * \varepsilon^{2} * \alpha_5 * \varepsilon^{1})$

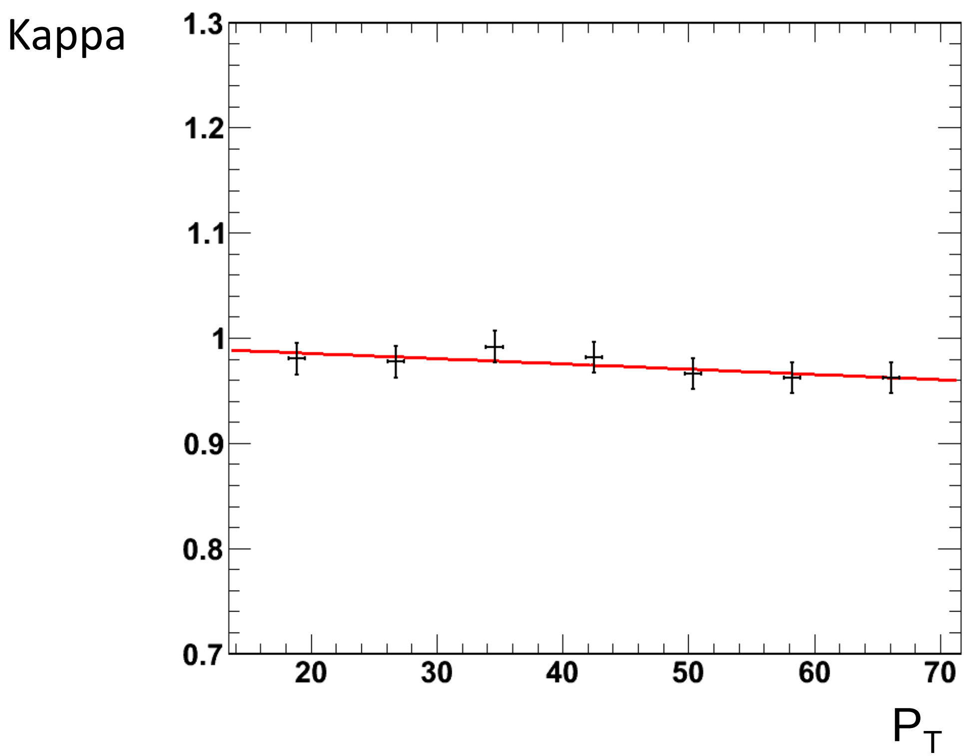
# Kappas determination (1/3)

→ We create some jet pT or Eta bins due to kappas' sensitivity to jet pT or Eta

→ We compute kappas for each bin

$$\kappa^{X,Y} = \varepsilon^{k_x,k_y} / (\varepsilon_{k_x} \times \varepsilon_{k_y})$$

→ We perform a linear fit to extract a function



# Kappas determination (2/3)

→ We must also compute kappas' errors

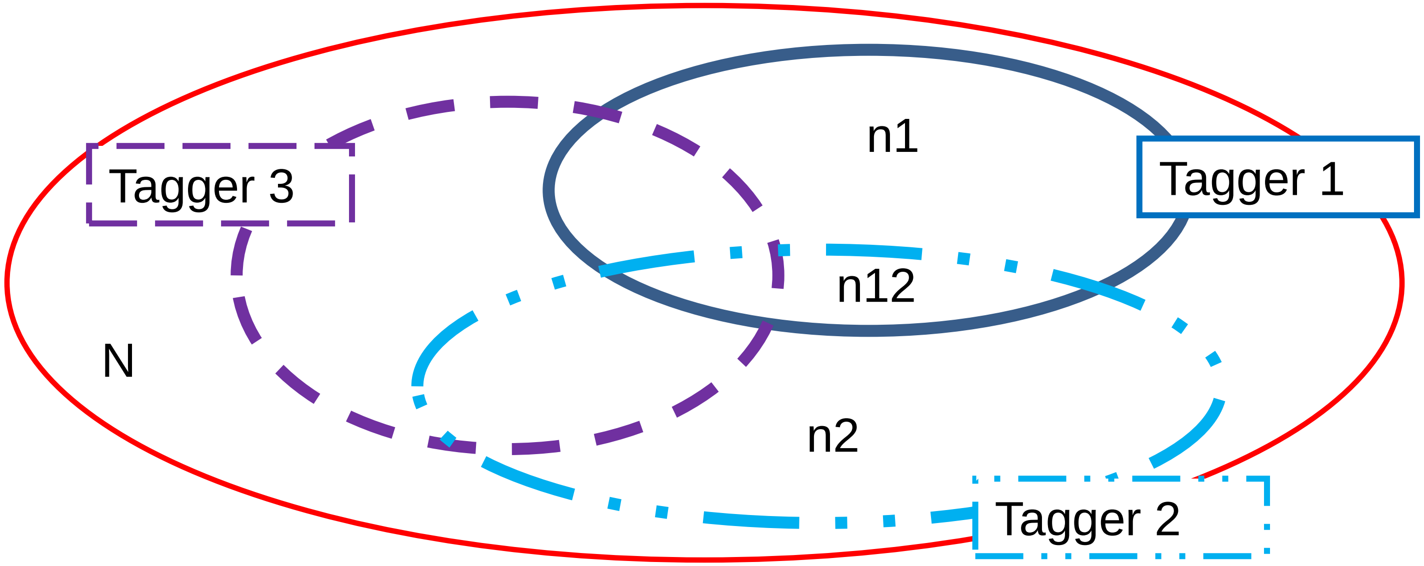
→ Suppose that the only source of uncertainty is statistical

→ We create independent samples : $w_i$ with $\sum_i w_i = n_X$

→ We shift $w_i$ → $w_i = w_i +$ Gaus(0 , 1 ) * sqrt($w_i$)

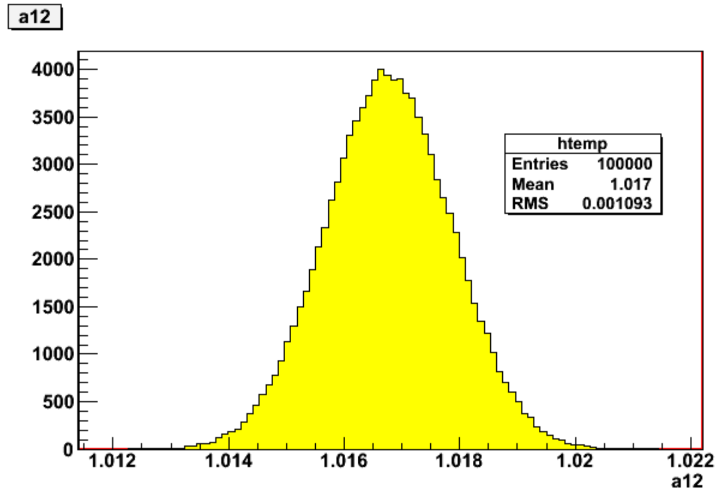
→ Recompute all kappas thousand times with shifted $w_i$

$$\kappa_{12} = \frac{n_{12} / N}{n_1 / N \times n_2 / N}$$

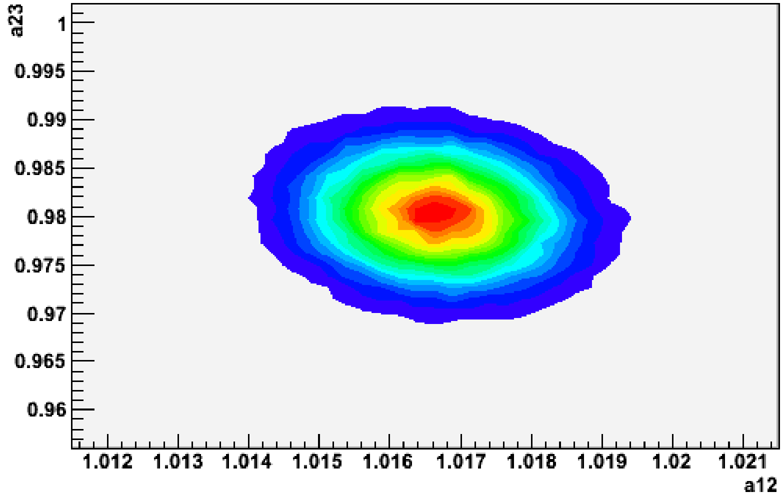


# Kappas determination (3/3)

Performing enough pseudo experiments we can find kappa's marginal PDFs




We do also have access to the non marginal kappa's distributions
→ We have covariance Matrices



# Error propagation for System8 (1/2)

**Benoit already talked about the analytic solver → Lets have a look to error propagation**

**Two sources of uncertainties :**

**→ Statistic**

**→ Systematic (error on kappas)**

Statistical source of uncertainty :

→ Limited data samples : $n_i$ follows poisson distribution

→ Standard approach (same as for kappas ) MC toy propagation

→ Creation of independent samples

→ Thousand of shifts and we solve the system for each iteration

→ Find the PDF for each unknowns

# Error propagation for System8 (2/2)

**Systematic are a bit more complicated**

**We can perform pseudo experiments shifting all kappas in there error bars using :**

**$K = K + \text{Gaus}(0 , 1) * \Delta K$**

**But kappas are (weakly) correlated → we can take into account this correlation using the covariance matrix**

→ Find the covariance matrix using pseudo exp in MC samples (slides 9-10)

→ We find eigen vectors and eigen values of the cov matrix

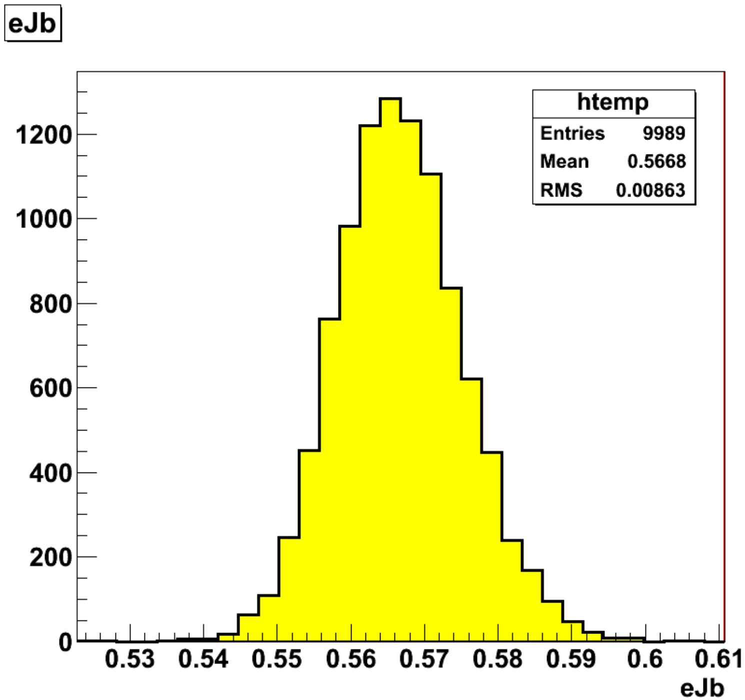
→ We created independent kappas (Place in the Eigen space of kappas)

→ Compute shifts for these kappas : **$\text{Gaus}(0 , 1) * \Delta K_{ind} = \text{Gaus}(0 , 1) * \text{sqrt(EigValue)}$**

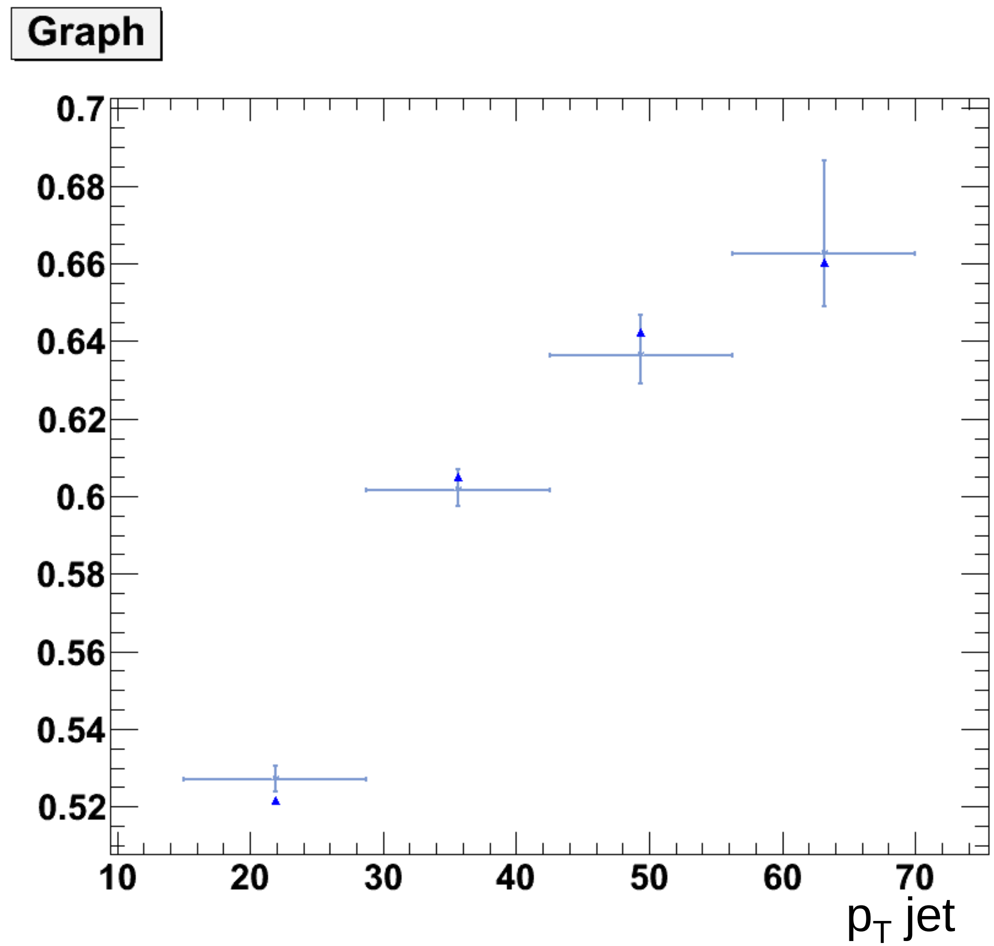
→ Recompute real kappas : **K = K + (EigVector)x(Kappa shift)**

→ Thousand of shifts and we solve the system for each iteration

# Preliminary Results




Efficiency distribution for SV0 on the topmix sample with $p_T$ jet ∈ [15;70] GeV after 10000 pseudo experiments




Binned efficiency for SV0 on the topmix sample with $p_T$ jet ∈ [15;70] with step 13 GeV

# The Soft Muon Tagger question …

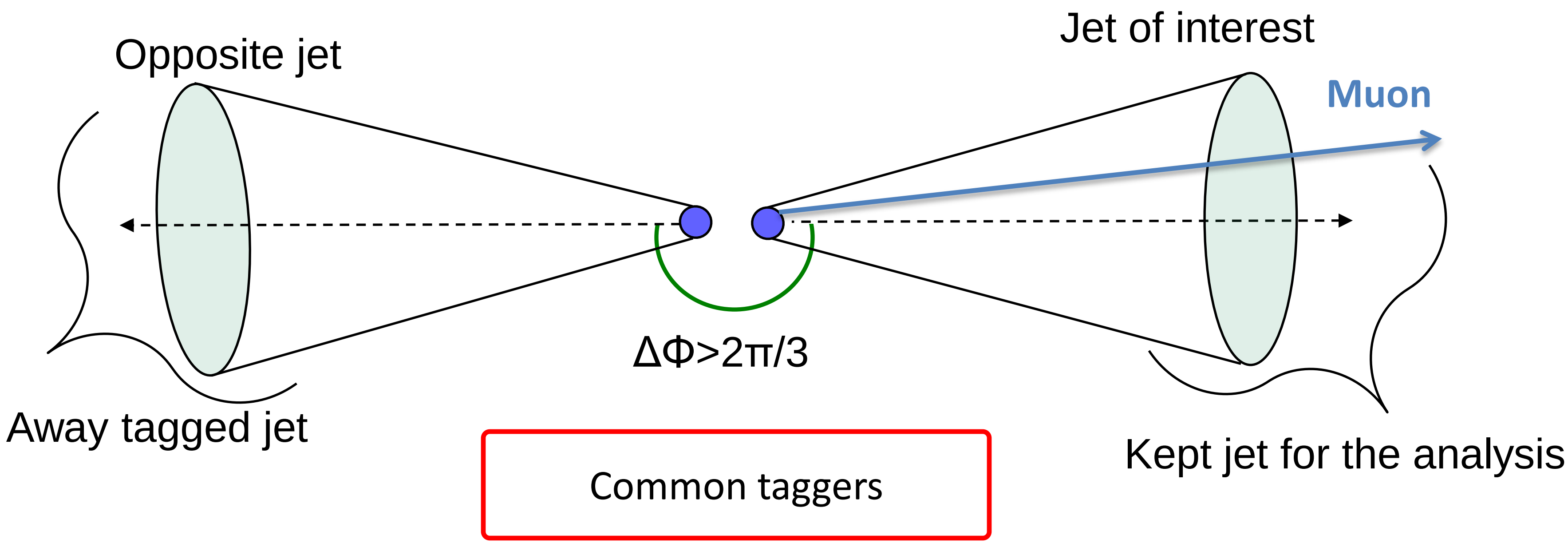
Is it relevant to use the soft muon tagger compared to $p_T^{rel}$?

Basically soft muon = $p_T^{rel}$ but also includes track selections
Can those track selections introduce a biais or correlation with the tagger of interest ??

Maybe safer to use just $p_T^{rel}$

# Idea to use system8 on all jets (1/2)

**One critical requirement from System8 is weakly correlated taggers**

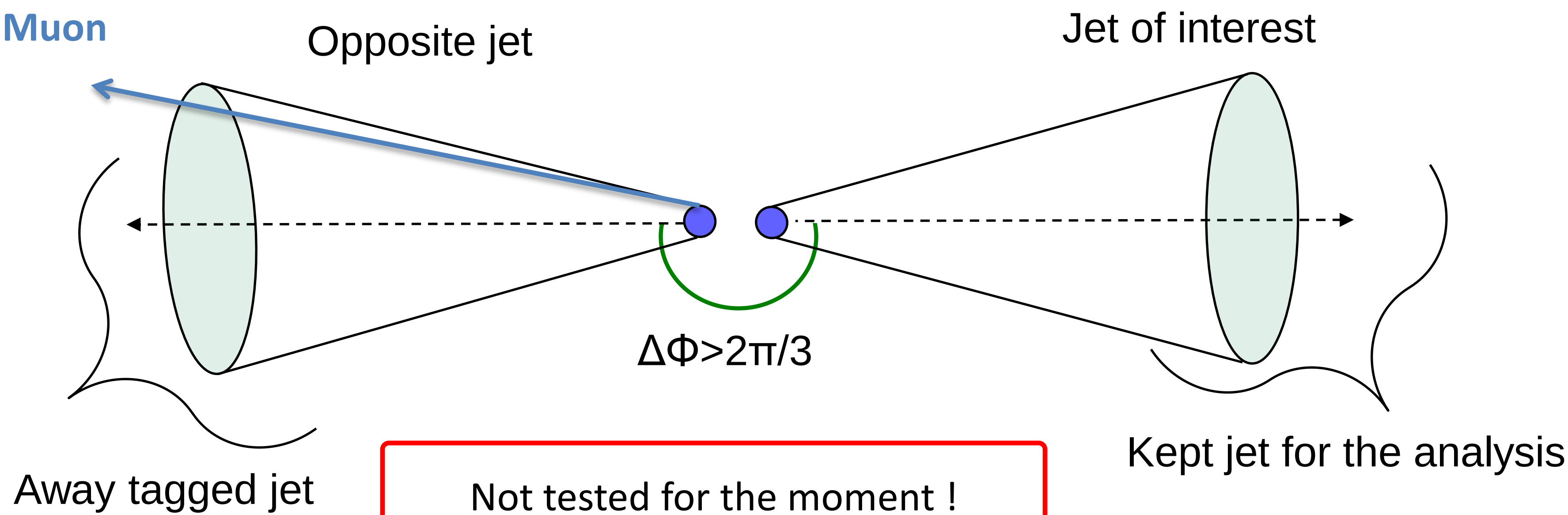



**Tagger 1 = the tagger of interest (IP3D, JETPROB, SV1, ...) is applied on the jet of interest**

**Tagger 2 = a soft muon tagger is applied on the jet of interest**

**Tagger 3 = presence of an opposite tagged jet**

# Idea to use system8 on all jets (2/2)




Not tested for the moment !

**Tagger 1 = the tagger of interest (IP3D, JETPROB, SV1, ...) is applied on the jet of interest**

**Tagger 2 = a soft muon tagger is applied on the "Away Jet"**

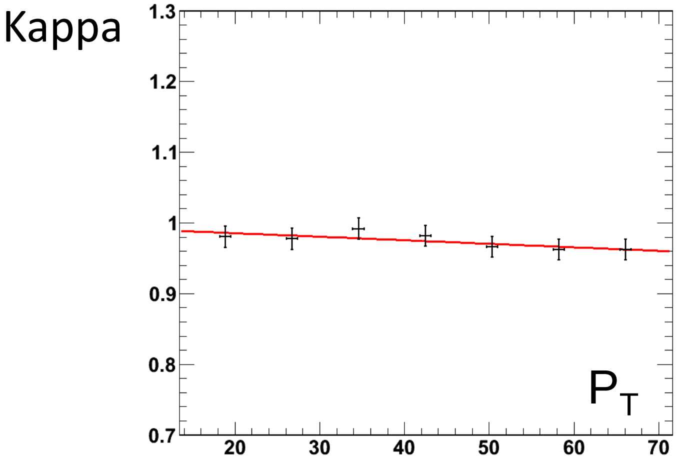
**Tagger 3 = an other tagger is applied on "Away Jet"**

# Conclusion

- Our system8 seems to work
- We define a strategy to compute kappas for system8
- We follow your recommendations to incorporate the cov matrix in System8
- We propagate errors in trought system8 using MC toy
- Still some rooms for improvement

# Kappas in System8

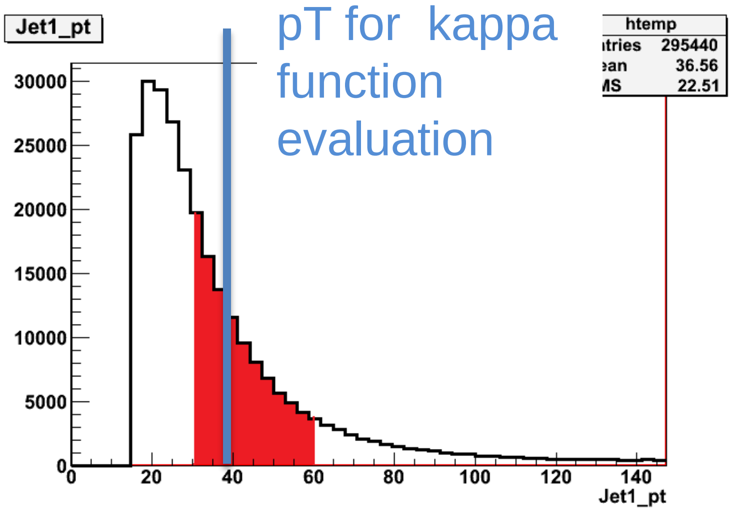
For a given value of pT or Eta we evaluate kappas on curves discussed before



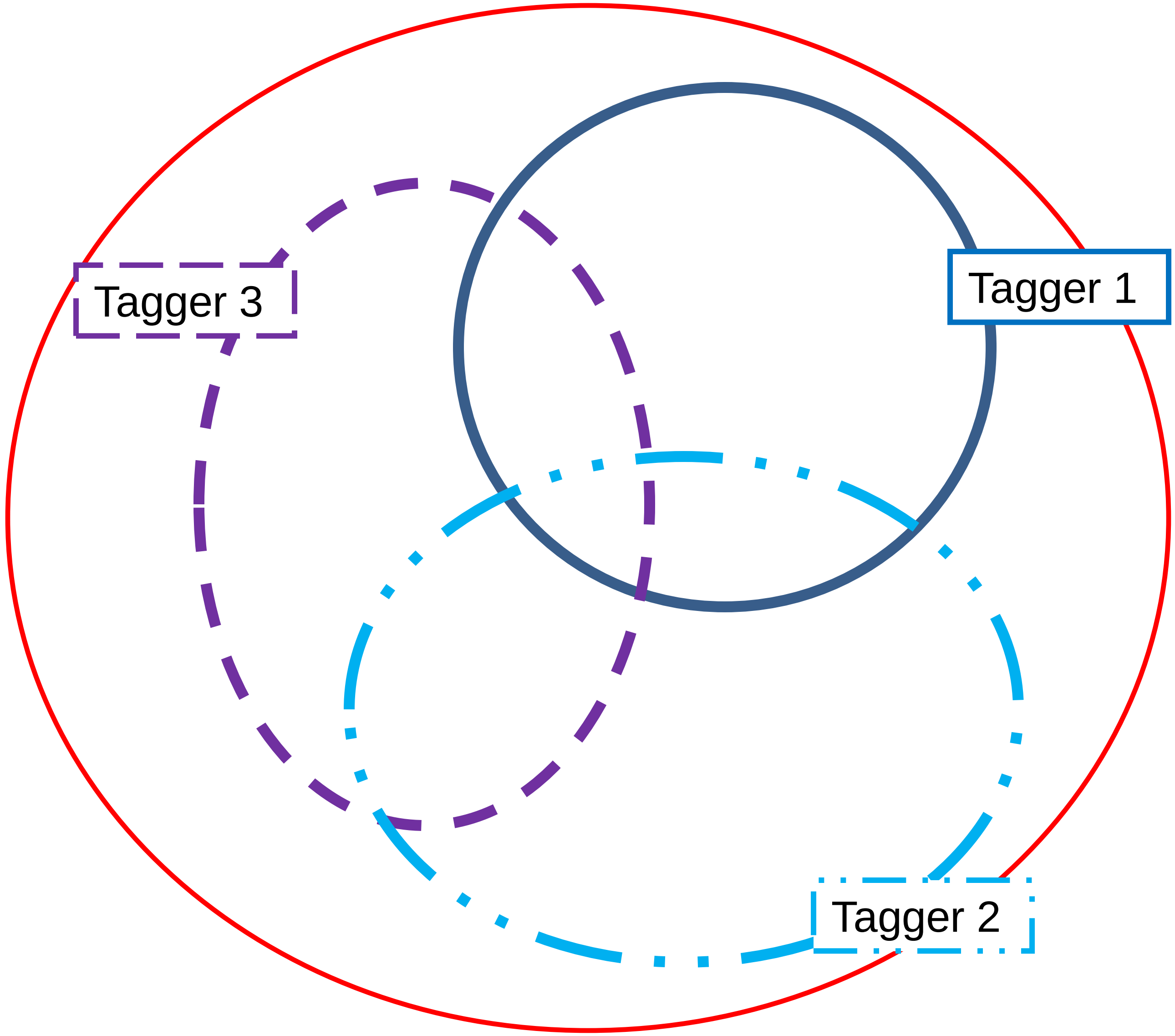

For a given range in pT which kappa should we take ?
f(pT=low) + f(pT= max) / 2
No cause this can introduce bias

We evaluate the kappa function for the pT that give :

$$p_T Eval \rightarrow \int_{lowbin}^{p_T Eval} f_\kappa(p_T)dp_T = \frac{\int_{lowbin}^{\max bin} f_\kappa(p_T)dp_T}{2}$$



# Graphical representation of system8



# Parameterisations

$$\varepsilon(p_T,\eta) = \frac{1}{\varepsilon^{All}}\left(\frac{c}{1+ae^{-bp_T}}\right) \times (d + e\eta + f\eta^2 + g\eta^3 + h\eta^4)$$