

# Status dell'analisi sugli algoritmi di b-tagging dell'esperimento CMS

Riunione CMS Torino Group

Alberto Traverso  
*Università degli Studi di Torino - INFN*

June 1, 2012

# OUTLINE

- ▶ Obiettivi del lavoro
- ▶ Definizione della principali variabili dell'analisi
- ▶ Test degli algoritmi di b-tagging
- ▶ Sviluppo di un tool per l'analisi della ricostruzione dei *jet*
- ▶ Misura della sezione d'urto di produzione associata  $Z + b$

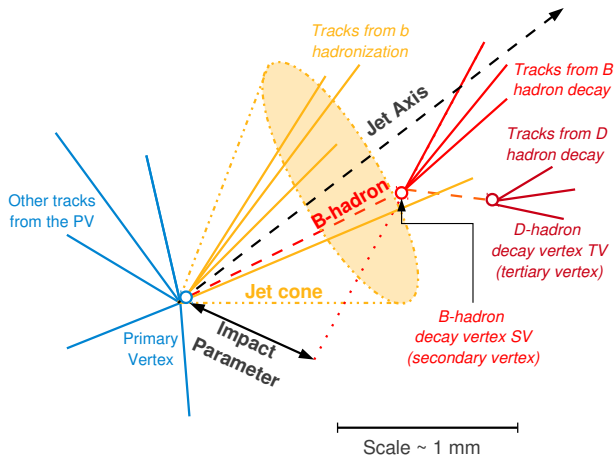
# OBIETTIVI DEL LAVORO

- ▶ Studiare e validare gli algoritmi di b-tagging di CMS qualora venisse cambiata la condizione di ricostruzione delle tracce
- ▶ Implementare un metodo di confronto della variabili *jet per jet* più sensibile del precedente
- ▶ Applicare le conoscenze ottenute in ambito degli algoritmi di b-tagging per misurare un processo fisico

# DEFINIZIONE DELLE PRINCIPALI VARIABILI DELL'ANALISI

- ▶ Poiché i quark pesanti (charm, bottom) hanno vita media relativamente lunga (decadono debolmente) e nella funzione di frammentazione gli adroni portano via circa il 70 % della energia essi volano attraverso il rivelatore
- ▶ Viene ricostruito un jet (in media 4 o 5 tracce cariche) combinando sia informazioni calorimetriche che di tracciatura.
- ▶ L'idea degli algoritmi di b-tagging è quella di associare un numero (*Float*) ad un jet ricostruito.
- ▶ In una situazione ideale si vorrebbe avere una distinzione netta tra jet prodotti da quark leggeri (bassi valori del discriminante) e quark pesanti (alti valori)

# DEFINIZIONE DELLE PRINCIPALI VARIABILI DELL'ANALISI



# DEFINIZIONE DELLE PRINCIPALI VARIABILI DELL'ANALISI

- ▶ A vertici secondari vengono associati alti valori del parametro di impatto (IP) o della significanza del parametro di impatto, ossia il parametro di impatto diviso il suo errore
- ▶ Algoritmi basati sul conteggio delle tracce (**T**rack **C**ounting) ad alto parametro di impatto: TCHE, TCHP,JP
- ▶ Algoritmi basati sulla ricostruzione di un vertice secondario (**S**imple **S**econdary **V**ertex): SSVHE,SSVHP
- ▶ Algoritmi che fanno uso di metodi statistici avanzati: CSV

## TEST SUGLI ALGORITMI DI B-TAGGING

- ▶ Scopo: introdurre geometrie misallineate nel ricostruire le tracce e studiare la sensibilità degli algoritmi di b-tagging rispetto a questi *bias*
- ▶ La prima verifica che è stata fatta è stato introdurre dei grossi spostamenti nei BPIX per testare la robustezza degli algoritmi, successivamente si sono considerati scenari più realistici
- ▶ Il tool utilizzato è quello ufficiale di CMS per la validazione del b-tagging
- ▶ Il campione utilizzato corrisponde ad 1 milione di eventi simulati  $t\bar{t}$ , il che garantisce l'esistenza di almeno due jet provenienti da b
- ▶ E' stato sviluppato anche un tool che permette di valutare le differenze di efficienza (rispetto ad una condizione ideale) a fissata efficienza dei jet leggeri (working point nelle analisi di CMS)

# ALCUNI RISULTATI SU CAMPIONI MC

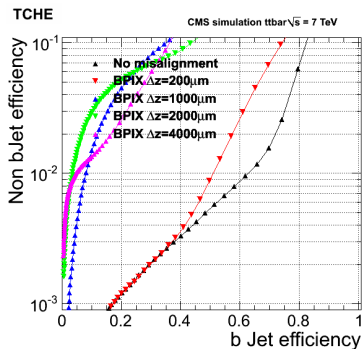


Figure: Efficienza dei light vs efficienza dei b per  $|\eta| < 2.4$  e  $p_T > 30\text{GeV}$  TCHE

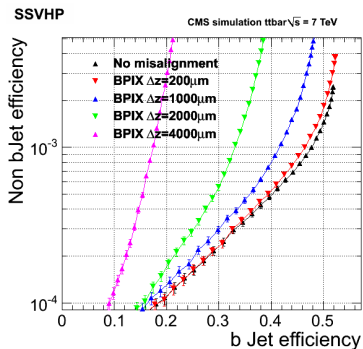


Figure: Efficienza dei light vs efficienza dei b per  $|\eta| < 2.4$  e  $p_T > 30\text{GeV}$  SSVHP



# ALCUNI RISULTATI SU CAMPIONI MC

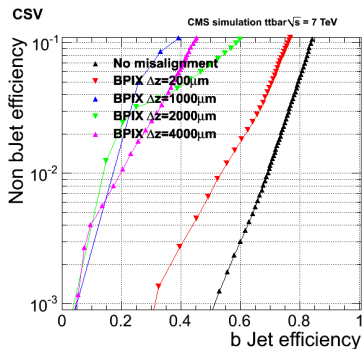


Figure: Efficienza dei light vs efficienza dei b per  $|\eta| < 2.4$  e  $p_T > 30\text{GeV}$  CSV

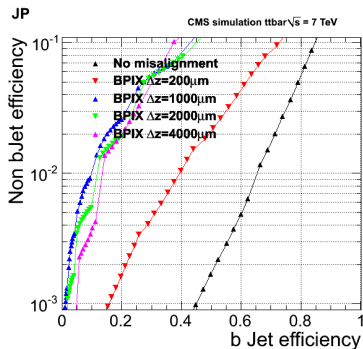


Figure: Efficienza dei light vs efficienza dei b per  $|\eta| < 2.4$  e  $p_T > 30\text{GeV}$  JP

# ALCUNI RISULTATI SU CAMPIONI MC

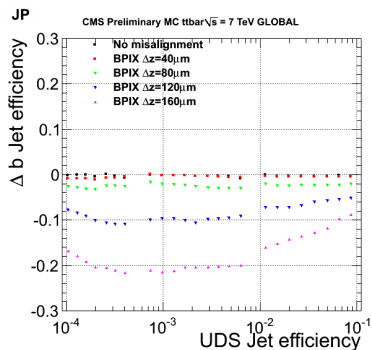


Figure: Differenze di efficienza dei b a fissata efficienza dei light per  $|\eta| < 2.4$  e  $p_T > 30\text{GeV}$  JP

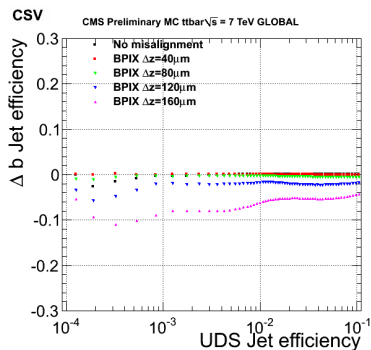
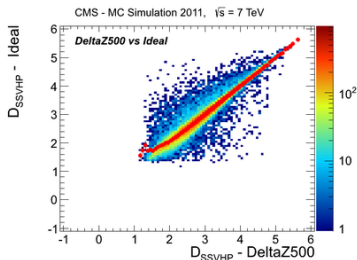
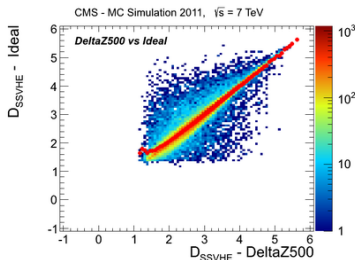
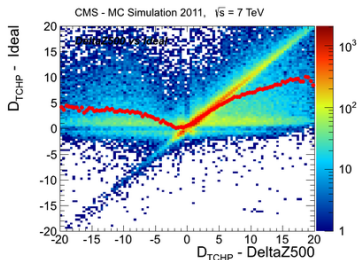
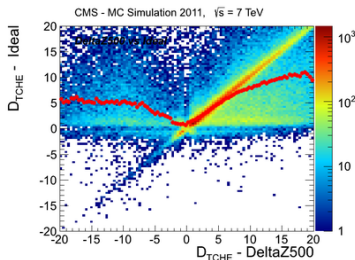


Figure: Differenze di efficienza dei b a fissata efficienza dei light per  $|\eta| < 2.4$  e  $p_T > 30\text{GeV}$  CSV

## CONSIDERAZIONI

- ▶ Le curve mostrate in precedenza nonostante segnalino una riduzione globale delle performances non sono molto sensibili a piccole variazioni
- ▶ Il software ufficiale di ricostruzione degli eventi di CMS non permette in generale di utilizzare due condizioni di ricostruzione diverse allo stesso tempo, ciò non dà la possibilità di valutare eventuali correlazioni
- ▶ Si è allora deciso di sviluppare un tool di diagnostica che permetta il confronto *jet per jet*, evento per evento tra due condizioni di ricostruzione differenti, con le seguenti caratteristiche principali:
- ▶ Valutazione di quantità non solo mediate, studio dei discriminanti in diverse regioni cinematiche,
- ▶ Produzione di una classe ad hoc (*JET info*) : distribuzioni di base dei jet ( $\eta, \phi, p_T$ ), dei discriminanti e del SV o PV
- ▶ Produzione di grafici divisi per flavor del jet: b (compreso il gluon splitting), light

# UN ESEMPIO DI RISULTATO



# CONSIDERAZIONI

- ▶ Dopo che il lavoro è stato presentato ufficialmente al CERN il tool è stato utilizzato da diversi gruppi (studi sul PU, fast simulation vs full simulation) (see James Keaveney)
- ▶ Ciò nasce da fatto che l'architettura è abbastanza versatile nei file in input: il tool permette il confronto di due scenari che differiscano per qualunque condizione iniziale
- ▶ In generale per quanto riguarda piccoli spostamenti non ci sono indicazioni di effetti netti a dimostrazione del fatto che le variabili studiate appaiono molto robuste, tuttavia lo strumento permette di effettuare una diagnostica molto raffinata

## MISURA DELLA SEZIONE D'URTO Z+B

- ▶ E' già uscito un articolo di CMS in cui si misura la sezione d'urto di produzione di Z+b con 2.1/fb di dati nel 2011 usando come discriminante il SSVHPT
- ▶ La sezione d'urto può essere così espressa:

$$\sigma_{hadron}(Z/\gamma^* + b, Z/\gamma^* \rightarrow \ell\ell) = \frac{N_{\ell\ell+b} \times (\mathcal{P} - f_{t\bar{t}})}{\mathcal{A}_\ell \times \mathcal{C}_{hadron} \times \epsilon_\ell \times \epsilon_b \times \mathcal{L}}$$

- ▶ Per il futuro si sta pensando di ridurre l'uso dei discriminanti limitatamente al CSV, in quanto attualmente risulta il più efficiente
- ▶ L'ipotesi originale era quella di fornire una misura utilizzando proprio il discriminante CSV, iniziando a stimare la purezza di b-tagging utilizzando come variabile *templata* per i fit la distribuzione della JP
- ▶ Ci si è trovati di fronte ad un problema tecnico in merito alla calibrazione della JP, siamo in attesa di presentare il problema al BTV

## STIMA DELLA PUREZZA

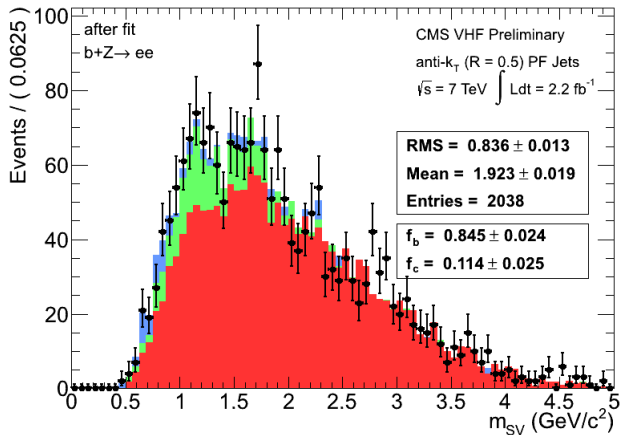


Figure: Fit ai template per la stima della purezza per il discriminante SSVHPT nel canale  $Z \rightarrow ee$

## MISURA DELLA SEZIONE D'URTO Z+B

- ▶ Il valore ottenuto nell'articolo (5352 eventi selezionati) è pari a :

$$5.84 \pm 0.08(stat) \pm 0.72(syst)_{-0.55}^{+0.25}(th.)pb$$

- ▶ In seguito ad una discussione interna al gruppo SMP-VJ è venuto fuori che il valore non è confrontabile con quello predetto da MCFM poiché il calcolo teorico considera solo il *leading* b-jet nell'evento
- ▶ Considerando questo topic molto urgente ed avendo a Torino buona parte del codice disponibile per valutare tutti i fattori della sezione d'urto sto terminando il calcolo della sezione d'urto compatibile con MCFM:

$$3.97 \pm 0.47pb$$