



# UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia “Galileo Galilei”

Corso di Laurea in Fisica

Tesi di Laurea

Studio per l'identificazione di kaoni con i dati di Belle II

Relatore

Prof. Roberto Stroili

Correlatrice

Dott.ssa Agnese Giaz

Laureanda

Giovanna Saleh

Anno Accademico 2019/2020



# Indice

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Introduzione</b>  | <b>1</b>  |
| 1.1      | Obiettivo . . . . .  | 1         |
| 1.2      | L'esperimento: SuperKEKB e Belle II . . . . .                    | 1         |
| 1.3      | Dataset . . . . .  | 4         |
| 1.4      | Metodi di analisi . . . . .                                      | 6         |
| <b>2</b> | <b>Efficienza relativa in funzione dell'angolo</b>               | <b>7</b>  |
| 2.1      | Metodo A . . . . .   | 7         |
| 2.2      | Metodo B . . . . .   | 8         |
| 2.3      | Metodo C . . . . .   | 10        |
| 2.3.1    | LOOSE . . . . .  | 10        |
| 2.3.2    | ALL . . . . .  | 12        |
| 2.4      | Osservazioni sulla scelta del metodo . . . . .                   | 14        |
| <b>3</b> | <b>Efficienza relativa complessiva</b>                           | <b>15</b> |
| <b>4</b> | <b>Efficienza assoluta in funzione delle selezioni sulla PID</b> | <b>17</b> |
|          | <b>Conclusioni</b>   | <b>19</b> |



# Capitolo 1

## Introduzione

### 1.1 Obiettivo

Obiettivo di questo lavoro di tesi è lo studio dell'efficienza del rivelatore TOP, attivo nella regione centrale dell'esperimento Belle II presso il collisore SuperKEKB, situato a Tsukuba, Giappone.

In particolare, viene preso in considerazione il decadimento  $\phi \rightarrow K^+K^-$ , che produce mesoni a basso momento e pertanto consente di analizzare l'efficienza dello strumento, nel quale arrivano le particelle aventi  $p \gtrsim 0.3$  GeV/c, anche nella regione con  $p < 0.5$  GeV/c, non coperta dai decadimenti analizzati negli altri studi della risposta del rivelatore attualmente in corso.

Si procede quindi allo studio dell'efficienza del TOP nell'identificazione di  $K^+$  e  $K^-$ .

A tale fine vengono analizzati i dati sperimentali raccolti da Belle II nel 2019, durante la fase 3<sup>1</sup> (proc11), per una luminosità totale di  $8.7 \text{ fb}^{-1}$ . I risultati vengono confrontati con quelli ottenuti effettuando la medesima analisi su dati generati sfruttando il metodo di simulazione Monte Carlo (MC). In particolare, i campioni di MC sono stati prodotti dalla simulazione MC13b run dependent utilizzando il programma di simulazione dell'esperimento, basato sul software GEANT4 [1]. La luminosità totale dei dati MC è di  $10 \text{ fb}^{-1}$ .

L'analisi dati viene svolta prevalentemente utilizzando il software ROOT [2] e in particolare il modulo RooFit [3].

### 1.2 L'esperimento: SuperKEKB e Belle II

Il Modello Standard, che ad oggi rappresenta la descrizione più completa del mondo subatomico, non spiega alcuni fenomeni, come ad esempio l'asimmetria tra materia e antimateria. Per rispondere alle domande ancora aperte, sono stati proposti diversi modelli e si stanno svolgendo numerosi esperimenti finalizzati alla scoperta di Nuova Fisica. La ricerca fondamentale nel campo della Nuova Fisica avviene adottando due approcci complementari [5]. Da un lato, è possibile lavorare ad alta energia (ATLAS, CMS) con l'obiettivo di scoprire nuove particelle prodotte in collisioni protone-protone. Dall'altro, è possibile lavorare ad alta intensità, con l'obiettivo di scoprire nuove interazioni deboli, oltre il Modello Standard, che potrebbero manifestarsi in stadi intermedi di processi rari. In questo secondo filone si inserisce l'esperimento Belle II.

Belle II, costituito da diversi sottorivelatori tra cui il TOP in esame, è installato nella regione di interazione del collisore SuperKEKB, upgrade del collisore KEKB, del centro di ricerca KEK, a Tsukuba, Giappone.

---

<sup>1</sup>Stato del rivelatore: installate tutte le componenti descritte nel paragrafo 1.2 eccetto una parte dei rivelatori a pixel

SuperKEKB, progettato per raggiungere una luminosità istantanea di  $8 \cdot 10^{35} \text{cm}^{-2}\text{s}^{-1}$ , fa collidere fasci ad energia asimmetrica di elettroni (7 GeV) e positroni (4 GeV). In questo modo si ottiene un boost lungo la direzione dei fasci nel sistema di riferimento del centro di massa, utile per consentire lo studio di violazioni di CP dipendenti dal tempo. L'energia delle collisioni, nel centro di massa, è circa pari a 10.5 GeV, ovvero prossima alla risonanza della  $\Upsilon(4S)$ : in tal modo l'energia è la minima necessaria per la produzione di coppie di mesoni B nel decadimento  $\Upsilon(4S) \rightarrow B\bar{B}$ . La probabilità che nel decadimento vengano prodotte anche altre particelle è quindi minima. Conseguentemente è minimo anche il fondo dovuto alla presenza di particelle non di interesse [6].

In SuperKEKB i fasci si muovono in due anelli distinti (fig. 1.1): gli elettroni nell'anello ad alta energia (HER), i positroni nell'anello a bassa energia (LER). I fasci collidono in un unico punto, in corrispondenza del sito dell'esperimento Belle II.

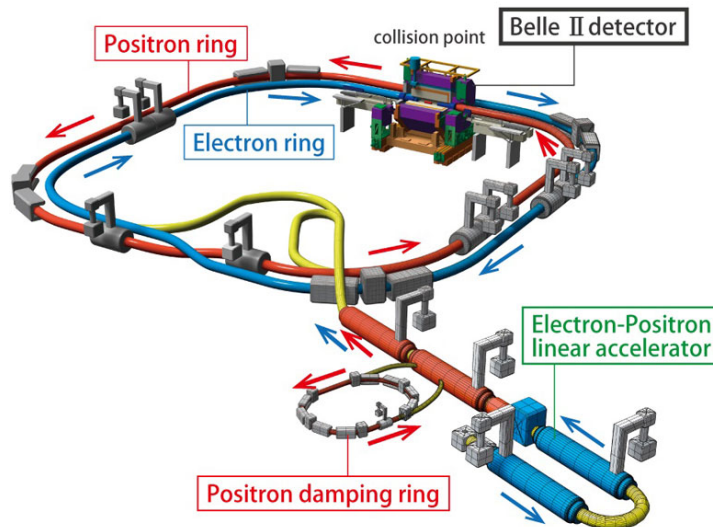


Figura 1.1: Schema collisore SuperKEKB. Immagine tratta da [6]

L'elevata luminosità è raggiunta implementando un modello di fascio (nanobeam) a bassissima emittanza, avente diametro di 50 nm in direzione  $y$  e  $5 \mu\text{m}$  in direzione  $x$  nel punto di interazione, combinato con un nuovo anello di smorzamento e una camera a vuoto per i positroni. [6]

A regime, ciascun fascio è costituito da 2500 pacchetti e le collisioni avvengono ogni 4 ns: con questi parametri, considerando le sezioni d'urto delle interazioni in esame, ci si aspetta vengano prodotte e rilevate circa  $10^3$  coppie di  $B\bar{B}$  al secondo, da cui il nome di B-factory all'acceleratore.

L'esperimento Belle II è situato lungo uno dei tratti rettilinei di SuperKEKB, ed è costituito da una successione concentrica di rivelatori deputati ciascuno ad una specifica misura (fig. 1.2).

Il punto di interazione (IP) è circondato da una beam pipe in berillio avente diametro molto ridotto (2 cm): ciò consente di posizionare due rivelatori al silicio a pixel (DEPFET) estremamente vicini all'IP, grazie ai quali si ottiene una precisa ricostruzione del vertice di decadimento. Questi sono circondati da altri quattro strati di rivelatori a bande di silicio a due facce. Le bande sono disposte ortogonalmente tra loro sui due lati del rivelatore, in modo da garantire una completa ricostruzione spaziale (coordinate  $xy$ ) della traccia.

È poi presente una camera a deriva (CDC) cilindrica riempita con una miscela di elio ed etano in uguale proporzione: il suo principale scopo è il tracciamento delle particelle che la attraversano, ma fornisce anche alcune informazioni utili alla loro identificazione, in particolare l'energia di ionizzazione rilasciata nel gas e il momento.

Il principale responsabile dell'identificazione delle particelle (particle identification, PID) è tuttavia il TOP (Time Of Propagation), un rivelatore Cherenkov costituito da barre di quarzo. Il TOP è composto da sedici moduli identici disposti attorno al punto d'interazione. Ciascun modulo è composto

da quattro parti [8]: due barre di quarzo, di dimensione ( $125 \times 45 \times 2$ ) cm, che agiscono come radiatore Cherenkov, uno specchio collocato ad una estremità della barra, avente la stessa sezione della barra stessa, e un prisma, collocato all'estremità opposta della barra, che collega la barra ad un array di fotomoltiplicatori (micro-channel-plate photomultiplier tubes, MCP-PMTs).

TOP misura il tempo di propagazione e il punto finale di impatto di un fotone Cherenkov. Nella parte in avanti del rivelatore è collocato ARICH (Aerogel based Ring Imaging Cherenkov detector), che misura il numero e la posizione dei fotoni Cherenkov incidenti. [5]

Segue poi il calorimetro elettromagnetico (ECL), utilizzato per rivelare i fotoni e identificare gli elettroni.

Tutta la strumentazione finora presentata è racchiusa all'interno di un magnete superconduttore solenoidale che genera un campo magnetico da 1.5 T.

Infine, esternamente al magnete, è collocato il rivelatore di  $K_L$  e muoni (KLM).

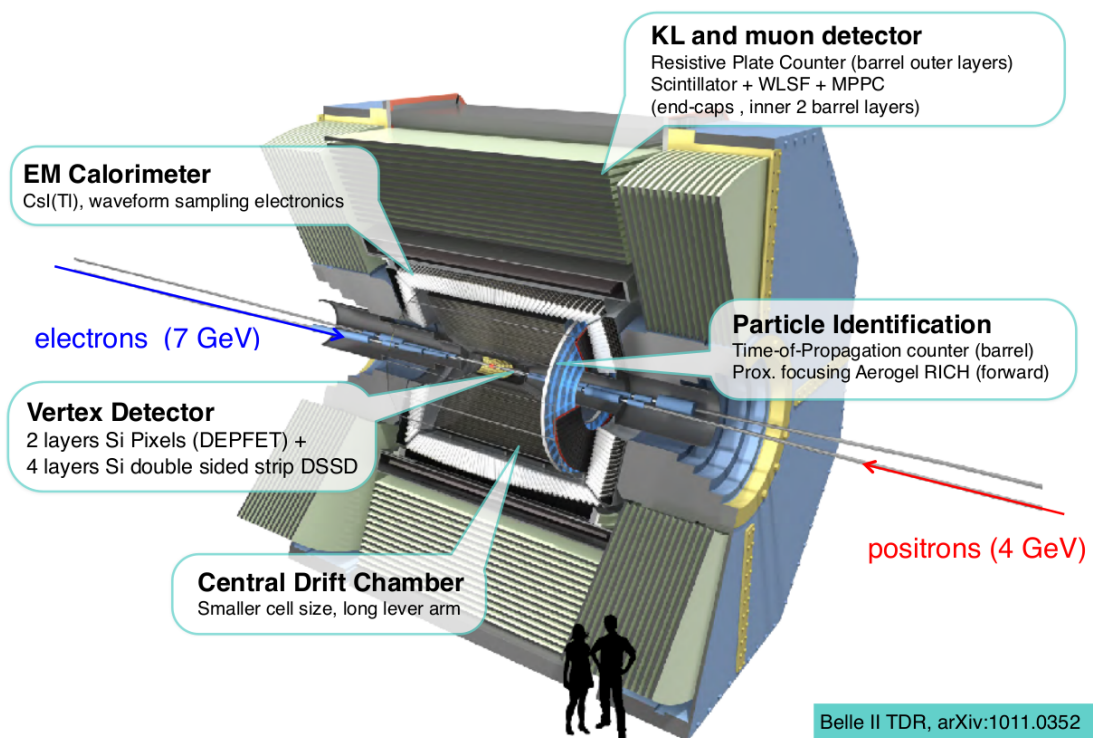


Figura 1.2: Schema esperimento Belle II. Immagine tratta da [5]

## 1.3 Dataset

In questo studio sono stati utilizzati i dati raccolti da Belle II nel run della Fase 3 (proc11, 2019). Per ciascun campione in esame, è stato considerato anche l'equivalente generato con la simulazione MC13b (in seguito MC), in modo da verificare l'adeguatezza della modellizzazione del rivelatore e mettere in luce eventuali scostamenti tra previsioni basate sul modello e dati sperimentali.

Tutta l'analisi viene svolta considerando due set di dati:

- **LOOSE**  $\rightarrow$  **KID** > 0.1

Nella fase preliminare di ricostruzione del decadimento, vengono selezionati solo gli eventi in cui il parametro legato alla Kaon Identification (KID) è maggiore di una soglia minima (0.1).

Si sottolinea che il parametro di KID viene stimato utilizzando informazioni provenienti da tutti i rivelatori di Belle II (e non solo dal TOP), pertanto applicare questo taglio nella creazione del dataset utilizzato per stimare l'efficienza del solo TOP a rigore costituisce un bias. Ci si aspetta quindi che l'efficienza calcolata con questi campioni sia leggermente più alta di quella reale del TOP. Si effettua comunque l'analisi sul dataset ottenuto con questo tipo di taglio, almeno nella prima fase dello studio, perché fornisce un campione pulito in cui il picco della massa invariante della  $\phi$  è molto evidente, il che consente di ottimizzare i successivi tagli sui dati e la metodologia di fit da utilizzare.

- **ALL**  $\rightarrow$  **No vincoli su KID**

Si considera il campione completo, senza alcun vincolo sulla KID. Applicando a questo dataset i tagli e i metodi di fit messi a punto grazie allo studio dei LOOSE, è possibile stimare l'efficienza effettiva del rivelatore TOP.

Per ricostruire il decadimento  $\phi \rightarrow K^+K^-$  vengono utilizzati i plot di massa invariante della  $\phi$ : per far emergere il segnale dal fondo, vengono effettuati i tagli riportati in tabella 1.1. Il risultato della selezione è esposto nei grafici 1.3, 1.4.

Tabella 1.1: Selezioni effettuate per la ricostruzione del decadimento  $\phi \rightarrow K^+K^-$

| Parametro  | Taglio  |
|--|---|
| Massa $\phi$   | $ m_\phi^{sper} - m_\phi^{PDG}  < 0.05 \text{ GeV}/c^2$ |
| Minima distanza tracce nel piano $r, \varphi$          | $ d0_{K^+} ,  d0_{K^-}  < 2 \text{ cm}$                 |
| Minima distanza tracce in direzione del fascio ( $z$ ) | $ z0_{K^+} ,  z0_{K^-}  < 4 \text{ cm}$                 |
| Numero CDC hits  | $nCDChits_{K^+}, nCDChits_{K^-} > 45$                   |
| Accettanza in momento                                  | $ptACC_{K^+}, ptACC_{K^-} = 1$                          |
| Accettanza geometrica                                  | $gACC_{K^+}, gACC_{K^-} = 1$                            |
| Accettanza angolare                                    | $thACC_{K^+}, thACC_{K^-} = 1$                          |
| Probabilità $\chi^2$ traccia                           | $p_{K^+}, p_{K^-} > 0.001$                              |



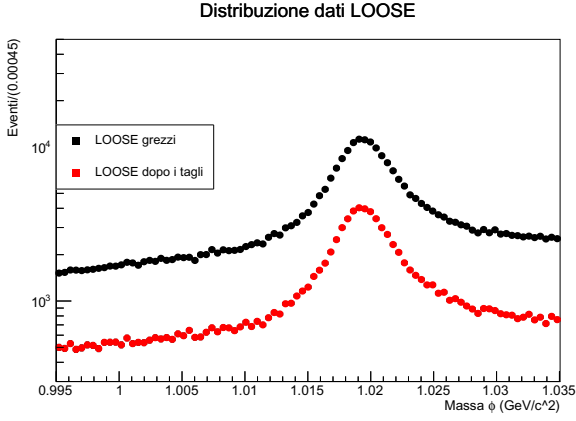


Figura 1.3: Tagli su dati LOOSE (C1)

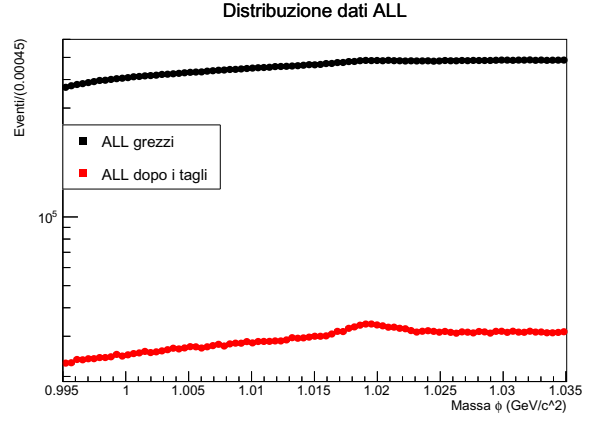


Figura 1.4: Tagli su dati ALL (C1)

Si sottolinea che nel caso degli ALL il picco emerso dalla selezione appare poco evidente nel grafico a causa della larghezza della scala in  $y$ , necessaria per esporre anche i dati grezzi. La forma effettiva di tale picco verrà discussa in modo più approfondito nelle successive analisi.

Nel corso dell'analisi si farà riferimento ai concetti di efficienza relativa e assoluta, intendendo la distinzione sulla base del tipo di vincolo applicato alla variabile di PID. In particolare, si parlerà di:

- **Efficienza relativa** → Vincolo applicato:  $\text{Prob}(\text{Kaone}) > \text{Prob}(\text{Pione})$ .  
Valutata l'efficienza nell'identificazione dei Kaoni rispetto ai Pioni.
- **Efficienza assoluta** → Vincolo applicato:  $\text{Prob}(\text{Kaone}) > \lambda$ , con  $\lambda$  soglia fissata a priori.  
Valutata l'efficienza nell'identificazione dei Kaoni.

Si intende innanzitutto valutare l'efficienza relativa del TOP nell'identificazione dei Kaoni rispetto ai Pioni. Si prendono quindi in esame le variabili di PID che contemplano solo le due possibilità Kaone/Pione per la particella osservata: PairK e PairPi, che restituiscono rispettivamente la probabilità che la particella sia un Kaone (e non un Pione) e viceversa, sulla base delle informazioni fornite dal solo rivelatore TOP. Nello specifico si definiscono:

$$\text{PairK} = \frac{\mathcal{L}_K}{\mathcal{L}_K + \mathcal{L}_\pi}$$

$$\text{PairPi} = \frac{\mathcal{L}_\pi}{\mathcal{L}_K + \mathcal{L}_\pi}$$

con  $\mathcal{L}$  massima verosimiglianza.

Per semplificare la notazione si definiscono i seguenti tre campioni, a cui si farà riferimento anche nel corso dell'analisi:

- **C1**: nessun vincolo sulla PID
- **C2**:  $K^- \text{PairK} > K^- \text{PairPi}$  e nessun vincolo su  $K^+ \text{ID}$
- **C3**:  $K^+ \text{PairK} > K^+ \text{PairPi}$  e nessun vincolo su  $K^- \text{ID}$

## 1.4 Metodi di analisi

L'efficienza di uno strumento nell'identificazione di una particella  $P$  è data dal rapporto tra numero di  $P$  rivelate e identificate come tali e numero di  $P$  incidenti:

$$\epsilon = \frac{\# P \text{ rivelate}}{\# P \text{ incidenti}}$$

Volendo valutare l'efficienza relativa del TOP nell'identificazione di  $K^+$  e  $K^-$ , si considerano:

$$\epsilon_{K^+} = \frac{\# C2}{\# C1} \qquad \epsilon_{K^-} = \frac{\# C3}{\# C1}$$

In questo studio vengono esaminati tre metodi per stimare quanti eventi sono contenuti in ciascun campione:

- **A** → Conteggio dei dati ( $\#$  dati) e fit della distribuzione della massa invariante con una funzione data dalla convoluzione di una gaussiana e una Breit-Wigner (segnale), sommata ad un polinomio di Chebyshev (fondo). La frazione di fondo (bkgfraction) è determinata nel fit come parametro.

$$\implies \#C_i = \#dati_i \cdot (1 - \text{bkgfraction}_i)$$

- **B** → Fit della distribuzione della massa invariante con una funzione data da una gaussiana di ampiezza  $A$  e varianza  $\sigma^2$  (segnale), sommata ad un polinomio di Chebyshev (fondo). Il numero di eventi è dato dall'integrale analitico della gaussiana.

$$\implies \#C_i = \sqrt{2\pi} A_i \sigma_i$$

- **C** → Fit della distribuzione della massa invariante con una funzione data dalla somma di due gaussiane (segnale), sommata ad un polinomio di Chebyshev (fondo). Il numero di eventi di segnale è determinato come parametro dal fit (nsig).

$$\implies \#C_i = \text{nsig}_i$$

Nelle successive sezioni si espongono i risultati dell'analisi effettuata con questi tre metodi e si procede alla determinazione del più efficace ai fini del calcolo dell'efficienza.

Per la verifica delle ipotesi sui fit si assume come confidence level  $C.L. = 0.001 = \int_{\xi}^{\infty} f(x) dx$ , in cui  $\xi$  rappresenta la soglia per cui l'integrale della densità di probabilità  $f(x)$  assume il valore prefissato, tenuto conto dei gradi di libertà relativi al campione in esame. Le ipotesi dei test in cui il p-value ( $p = \int_{\xi_{\text{spere}}}^{\infty} f(x) dx$ ), ossia la probabilità di ottenere un valore uguale o più estremo di quello ottenuto ( $\xi_{\text{spere}}$ ) risultasse  $p \leq C.L.$ , verranno rifiutate.

La scelta del grado del polinomio con cui modellizzare il fondo è stata fatta sulla base dei risultati del test del chi quadro su ciascun fit: viene sempre utilizzato il polinomio di grado minimo che garantisca  $p > C.L.$

Tutti gli errori verranno forniti con una cifra significativa, eccetto quando la prima è 1: in tal caso verrà esposta anche la seconda, affinché l'arrotondamento effettuato comporti una variazione pari al più al 25% dell'errore stesso.

## Capitolo 2

# Efficienza relativa in funzione dell'angolo

Il rivelatore TOP copre l'intervallo angolare  $\theta \in [30^\circ - 120^\circ]$ , con  $\theta$  angolo polare.

Per la successiva analisi si suddivide tale regione in intervalli di ampiezza pari a  $\Delta\theta = 10^\circ$  e si valuta l'efficienza relativa del TOP nell'identificazione di  $K^+$  e  $K^-$  in ciascun intervallo.

In prima approssimazione ci si aspetta che tale efficienza sia pressoché costante nei vari intervalli. Tuttavia, poiché il numero di fotoni Cherenkov rivelati dal TOP varia con l'angolo, sono possibili delle deviazioni da questo comportamento.

### 2.1 Metodo A

La funzione utilizzata per il fit è quella che, secondo il modello teorico, meglio descrive i dati. La Breit-Wigner rappresenta la risonanza della  $\phi$ , avente larghezza  $\Gamma_\phi^{pdg} = 4.249 \pm 0.013$  MeV [9], mentre la gaussiana tiene conto degli effetti di risoluzione sperimentale.

Osservando la forma del fondo, si ritiene che un polinomio di Chebyshev di primo grado sia sufficiente per descriverlo. I valori ottenuti per il  $\chi^2$  consentono di accettare questa ipotesi al C.L. fissato per i campioni LOOSE. Per gli ALL risulta invece necessario considerare un polinomio di secondo grado.

Il fit viene eseguito con il software RooFit; si espone, a titolo di esempio, il grafico di un fit eseguito con questo metodo (fig. 2.1).

Come precedentemente indicato, è possibile stimare il numero di eventi in ciascun campione come  $\#C_i = \#\text{dati}_i \cdot (1 - \text{bkgfraction}_i)$ , in cui la frazione di fondo è un parametro fornito dal fit.

Si espongono di seguito i risultati ottenuti per le efficienze.

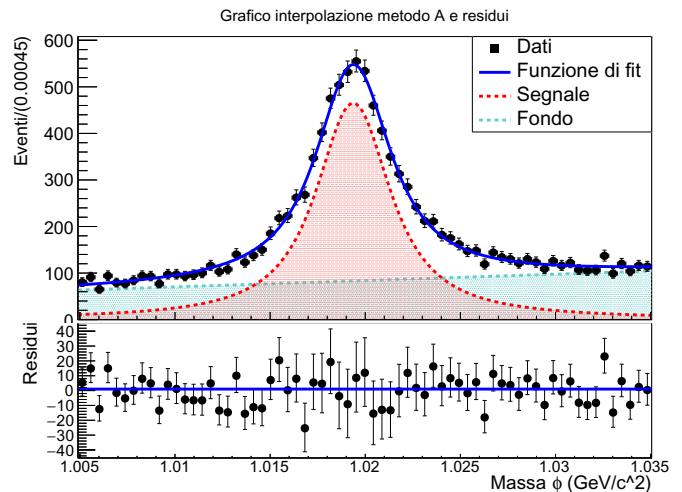


Figura 2.1: *Sopra: picco di massa invariante della  $\phi$  nell'intervallo  $60^\circ - 70^\circ$ , dati sperimentali, taglio LOOSE, campione C1, vincolo su  $\theta_{K^-}$ . In blu il fit totale, in rosso il segnale, in azzurro il fondo. Sotto: residui.*

Tabella 2.1: Risultati efficienza - Metodo A

| $\Delta\theta(^{\circ})$ | LOOSE           |                   |                 |                 | ALL             |                 |                 |                 |
|--------------------------|-----------------|-------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                          | Eff. Id. $K^+$  |                   | Eff. Id. $K^-$  |                 | Eff. Id. $K^+$  |                 | Eff. Id. $K^-$  |                 |
|                          | Dati            | MC                | Dati            | MC              | Dati            | MC              | Dati            | MC              |
| 30-40                    | $0.86 \pm 0.07$ | $0.81 \pm 0.03$   | $0.78 \pm 0.06$ | $0.82 \pm 0.03$ | $0.56 \pm 0.07$ | $0.70 \pm 0.09$ | $0.56 \pm 0.08$ | $0.65 \pm 0.06$ |
| 40-50                    | $0.81 \pm 0.06$ | $0.855 \pm 0.018$ | $0.89 \pm 0.09$ | $0.86 \pm 0.03$ | $0.75 \pm 0.09$ | $0.73 \pm 0.05$ | $0.68 \pm 0.07$ | $0.86 \pm 0.07$ |
| 50-60                    | $0.89 \pm 0.06$ | $0.85 \pm 0.03$   | $0.84 \pm 0.05$ | $0.87 \pm 0.02$ | $0.99 \pm 0.15$ | $0.76 \pm 0.05$ | $0.86 \pm 0.14$ | $0.75 \pm 0.07$ |
| 60-70                    | $0.83 \pm 0.04$ | $0.87 \pm 0.02$   | $0.84 \pm 0.06$ | $0.87 \pm 0.02$ | $0.58 \pm 0.13$ | $0.62 \pm 0.04$ | $0.61 \pm 0.06$ | $0.55 \pm 0.04$ |
| 70-80                    | $0.83 \pm 0.04$ | $0.87 \pm 0.05$   | $0.87 \pm 0.06$ | $0.87 \pm 0.03$ | $0.59 \pm 0.09$ | $0.76 \pm 0.05$ | $0.56 \pm 0.11$ | $0.74 \pm 0.06$ |
| 80-90                    | $0.80 \pm 0.08$ | $0.85 \pm 0.03$   | $0.83 \pm 0.09$ | $0.85 \pm 0.03$ | $0.69 \pm 0.11$ | $0.74 \pm 0.03$ | $0.60 \pm 0.13$ | $0.69 \pm 0.07$ |
| 90-100                   | $0.83 \pm 0.09$ | $0.83 \pm 0.03$   | $0.80 \pm 0.04$ | $0.84 \pm 0.03$ | $0.74 \pm 0.13$ | $0.6 \pm 0.1$   | $0.6 \pm 0.1$   | $0.81 \pm 0.04$ |
| 100-110                  | $0.7 \pm 0.1$   | $0.81 \pm 0.02$   | $0.8 \pm 0.1$   | $0.81 \pm 0.03$ | $0.56 \pm 0.13$ | $0.78 \pm 0.04$ | $0.76 \pm 0.17$ | $0.73 \pm 0.08$ |
| 110-120                  | $0.75 \pm 0.11$ | $0.72 \pm 0.03$   | $0.69 \pm 0.04$ | $0.73 \pm 0.03$ | $0.38 \pm 0.07$ | $0.62 \pm 0.13$ | $0.52 \pm 0.09$ | $0.65 \pm 0.04$ |

I risultati dell'analisi sui dati LOOSE sono conformi alle aspettative (formulate sulla base dei risultati di precedenti studi dell'efficienza [4]), mentre quelli relativi ai dati ALL non sono ottimali. In particolare il problema, emerso durante l'esecuzione dei fit e l'analisi dei risultati, è che il valore del parametro  $bkgfraction$  (l'unico degli otto forniti dal fit effettivamente utilizzato nel calcolo dell'efficienza), è estremamente sensibile ai valori iniziali forniti. Ciò, unito al fatto che non si ha una stima teorica della frazione di fondo, motivo per cui non è possibile ritenere un valore iniziale migliore di un altro, rende impossibile stabilire quale sia la stima più accurata di  $bkgfraction$  tra le varie ottenibili in uno stesso campione. L'errore sul  $bkgfraction$ , che giustamente tiene conto della grande incertezza sul valore, si traduce in una bassa precisione nella stima delle efficienze, come si osserva dai risultati esposti in tabella 2.1.

Nel caso dei dati ALL si ottengono quindi delle stime poco accurate e poco precise, motivo per cui si ritiene che questo metodo, seppur teoricamente corretto, non sia il più adatto allo scopo di stimare l'efficienza del TOP, in quanto i risultati del fit non sono soddisfacenti.

Una conferma dell'inadeguatezza dei risultati del fit deriva dalle stime fornite per la larghezza della Gaussiana e della Breit Wigner, mediamente pari a  $\sigma_{gaus}^{fit} \approx 0.004$  GeV e  $\gamma_{BW}^{fit} < 0.0001$  GeV. In particolare, confrontando questi risultati con il valore atteso di  $\Gamma_{\phi}^{pdg} = 4.249 \pm 0.013$  MeV, si evince che il fit converge utilizzando quasi solo la gaussiana per descrivere la forma del segnale. Questo risultato vanifica l'utilizzo di una funzione complessa come è la convoluzione delle due densità di probabilità in esame, e con esso l'utilizzo del metodo **A**.

## 2.2 Metodo B

I risultati ottenuti col metodo **A** portano a ritenere che sia possibile trascurare il contributo dato dalla Breit-Wigner, e dunque interpolare la distribuzione della massa invariante solamente con una gaussiana, sempre sommata ad un fondo modellizzato come polinomio di Chebyshev. Il fit viene eseguito con il software Root.

Il vantaggio di questo approccio è che la gaussiana è una funzione integrabile analiticamente, quindi non è necessario inserire il rapporto segnale/rumore tra i parametri di fit. Definita  $A$  l'ampiezza della gaussiana e  $\sigma^2$  la sua varianza, fornite dal fit, l'integrale, ovvero il numero di eventi di segnale nel campione, è dato da  $\#C_i = \sqrt{2\pi}A_i\sigma_i$ . L'errore su tale valore è stato calcolato considerando il contributo di covarianza tra i due parametri  $A, \sigma$  della gaussiana.

Le efficienze ottenute stimando in questo modo il numero di eventi sono riportate in tabella 2.2.

Tabella 2.2: Risultati efficienza - Metodo B

| $\Delta\theta(^{\circ})$ | LOOSE           |                   |                 |                   | ALL             |                 |                 |                 |
|--------------------------|-----------------|-------------------|-----------------|-------------------|-----------------|-----------------|-----------------|-----------------|
|                          | Eff. Id. $K^+$  |                   | Eff. Id. $K^-$  |                   | Eff. Id. $K^+$  |                 | Eff. Id. $K^-$  |                 |
|                          | Dati            | MC                | Dati            | MC                | Dati            | MC              | Dati            | MC              |
| 30-40                    | $0.83 \pm 0.03$ | $0.82 \pm 0.02$   | $0.80 \pm 0.03$ | $0.812 \pm 0.019$ | $0.90 \pm 0.06$ | $0.79 \pm 0.04$ | $0.77 \pm 0.09$ | $0.97 \pm 0.05$ |
| 40-50                    | $0.85 \pm 0.03$ | $0.814 \pm 0.014$ | $0.85 \pm 0.03$ | $0.862 \pm 0.015$ | $0.93 \pm 0.04$ | $0.98 \pm 0.04$ | $0.88 \pm 0.08$ | $0.79 \pm 0.03$ |
| 50-60                    | $0.87 \pm 0.03$ | $0.828 \pm 0.014$ | $0.86 \pm 0.03$ | $0.860 \pm 0.015$ | $0.88 \pm 0.06$ | $0.81 \pm 0.03$ | $0.76 \pm 0.08$ | $0.76 \pm 0.03$ |
| 60-70                    | $0.76 \pm 0.02$ | $0.827 \pm 0.015$ | $0.86 \pm 0.03$ | $0.875 \pm 0.016$ | $0.73 \pm 0.07$ | $0.80 \pm 0.03$ | $0.82 \pm 0.05$ | $0.83 \pm 0.03$ |
| 70-80                    | $0.86 \pm 0.03$ | $0.823 \pm 0.016$ | $0.92 \pm 0.03$ | $0.862 \pm 0.016$ | $0.73 \pm 0.04$ | $0.80 \pm 0.04$ | $0.77 \pm 0.04$ | $0.81 \pm 0.04$ |
| 80-90                    | $0.85 \pm 0.03$ | $0.801 \pm 0.016$ | $0.90 \pm 0.04$ | $0.850 \pm 0.017$ | $0.66 \pm 0.05$ | $0.81 \pm 0.04$ | $0.74 \pm 0.05$ | $0.83 \pm 0.04$ |
| 90-100                   | $0.83 \pm 0.03$ | $0.774 \pm 0.016$ | $0.75 \pm 0.03$ | $0.887 \pm 0.019$ | $0.78 \pm 0.06$ | $0.84 \pm 0.04$ | $0.80 \pm 0.06$ | $0.84 \pm 0.04$ |
| 100-110                  | $0.78 \pm 0.03$ | $0.819 \pm 0.018$ | $0.77 \pm 0.03$ | $0.782 \pm 0.018$ | $0.89 \pm 0.08$ | $0.80 \pm 0.04$ | $0.86 \pm 0.05$ | $0.84 \pm 0.04$ |
| 110-120                  | $0.73 \pm 0.04$ | $0.708 \pm 0.019$ | $0.76 \pm 0.04$ | $0.76 \pm 0.02$   | $0.83 \pm 0.09$ | $0.68 \pm 0.04$ | $0.75 \pm 0.07$ | $0.59 \pm 0.04$ |

Questo metodo risulta piuttosto efficace nella trattazione dei dati LOOSE: la funzione di fit descrive adeguatamente i dati, come confermato dai valori assunti dal  $\chi^2$ , che consentono per tutti i campioni di accettare i risultati ottenuti a C.L. fissato. L'errore associato ai parametri di fit utilizzati per il calcolo dell'integrale è abbastanza contenuto e il contributo dato dal termine di covarianza ( $\text{cov}(A, \sigma)$ ) è sempre negativo, pertanto riduce ulteriormente l'errore sul valore finale dell'integrale: ciò porta ad un errore compreso tra il 3% e il 5% sui valori delle efficienze. Si riscontra inoltre un discreto accordo tra i risultati ottenuti analizzando i dati sperimentali e i MC: in quasi tutti gli intervalli angolari l'efficienza sperimentale è compatibile entro i  $2\sigma$  con quella derivante dalla simulazione.

Per i dati ALL, invece, il modello utilizzato in questo metodo non risulta ottimale: la funzione di fit non descrive adeguatamente i dati, come si riesce in parte ad evincere già dall'osservazione dei grafici (esempio in fig. 2.2) e come confermato dai valori del  $\chi^2$ , che non consentono di accettare i risultati dei fit al C.L. fissato, nemmeno aumentando il grado del polinomio utilizzato per il fondo.

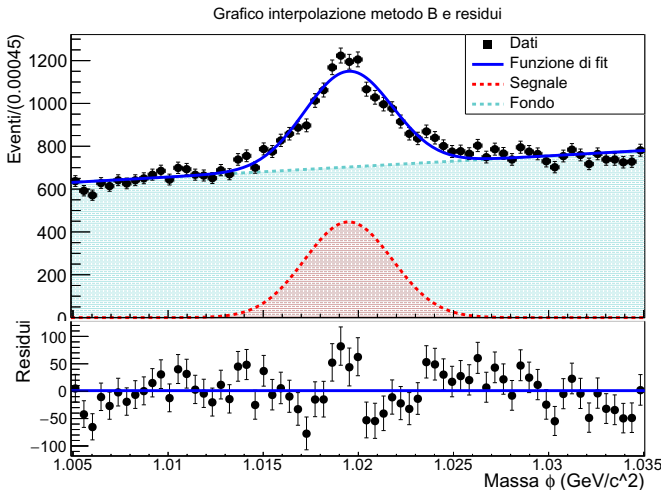


Figura 2.2: Sopra: picco di massa invariante della  $\phi$  nell'intervallo  $50^{\circ} - 60^{\circ}$ , dati sperimentali, taglio ALL, campione C2, vincolo su  $\theta_{K^+}$ . In blu il fit totale, in rosso il segnale, in azzurro il fondo. Sotto: residui.

L'ampiezza della gaussiana tende a sottostimare lievemente, ma sistematicamente, l'altezza effettiva del picco e le spalle della curva risultano spesso più larghe della distribuzione dei dati.

Questo effetto non era stato riscontrato nelle interpolazioni effettuate col metodo A: è quindi possibile che il contributo dato dalla Breit Wigner di larghezza estremamente ridotta portasse ad una funzione complessiva più piccata della semplice gaussiana di larghezza  $\sigma \approx 0.004$  GeV.

I risultati finali ottenuti per le efficienze con il metodo B appaiono piuttosto buoni, ma si ritiene che ciò sia il frutto di un'opportuna compensazione di sovrastima e sottostima della distribuzione in porzioni diverse del dominio, più che di un'effettiva aderenza del modello ai dati.

Inoltre, in alcuni campioni le barre d'errore consentirebbero di ottenere il risultato non fisico di un'efficienza  $\epsilon > 1$ : sarebbe possibile limitare questi effetti effettuando una trattazione dell'errore che introduca una distribuzione asimmetrica attorno al centroide e identicamente nulla per  $\epsilon > 1$ , come illustrato in [7].

Poiché tuttavia si ritiene che in questo caso il problema sia un'inadeguatezza di fondo del modello, forse troppo semplificato per descrivere l'effettiva distribuzione dei dati, si sceglie di non effettuare questo tipo di correzioni e procedere con il metodo successivo.

## 2.3 Metodo C

È possibile raffinare il metodo precedentemente esposto modellizzando il segnale come somma di due gaussiane, sempre sommate ad un fondo descritto da un polinomio di Chebyshev. In questo caso il fit, essendo più complesso, viene eseguito con l'estensione RooFit, in cui tuttavia la funzione viene definita come densità di probabilità e pertanto risulta automaticamente normalizzata sul dominio. Non ha quindi senso, ai fini dell'analisi che si intende svolgere, calcolare l'integrale analitico della distribuzione che descrive il segnale, poiché risulterebbe identicamente uguale a 1 in tutti i campioni, indipendentemente dal numero di eventi presenti in ciascuno di essi.

Si procede quindi con un Extended Maximum Likelihood fit [3], che consente stimare come parametro anche il numero di eventi di segnale ( $nsig$ ) e di rumore ( $nbg$ ). Il numero di eventi di segnale di ciascun campione è quindi dato direttamente da  $\#C_i = nsig_i$ .

### 2.3.1 LOOSE

Come si osserva dai grafici esposti a titolo di esempio (fig. 2.3, 2.4), e come confermato dai valori del  $\chi^2$ , nel caso dei campioni con taglio LOOSE è sufficiente un polinomio di primo grado per fittare adeguatamente il fondo. Si osserva inoltre che la funzione usata per interpolare il segnale descrive bene l'andamento dei dati, sia nella regione del punto di massimo del picco, sia lungo le spalle della distribuzione.

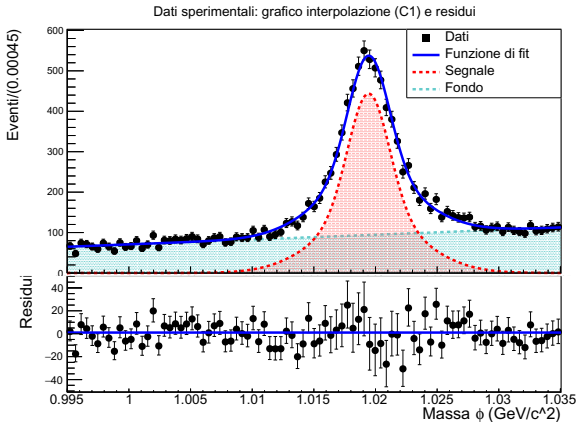


Figura 2.3: Sopra: picco di massa invariante della  $\phi$  nell'intervallo  $60^\circ - 70^\circ$ , dati sperimentali, taglio LOOSE, campione C1, vincolo su  $\theta_{K^+}$ . In blu il fit totale, in rosso il segnale, in azzurro il fondo. Sotto: residui.

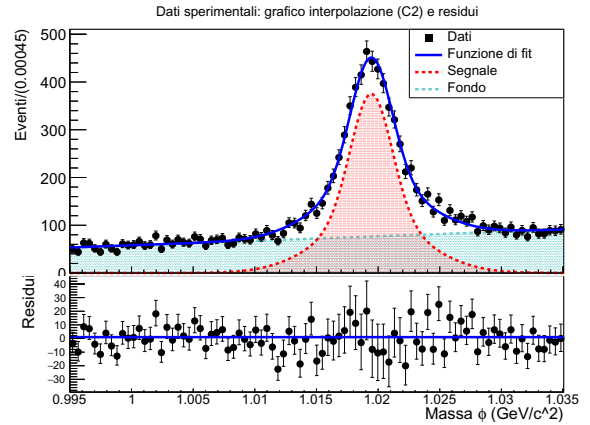


Figura 2.4: Sopra: picco di massa invariante della  $\phi$  nell'intervallo  $60^\circ - 70^\circ$ , dati sperimentali, taglio LOOSE, campione C2, vincolo su  $\theta_{K^+}$ . In blu il fit totale, in rosso il segnale, in azzurro il fondo. Sotto: residui.

Le stime delle efficienze ottenute utilizzando questo modello per calcolare il numero di eventi sono esposte in tabella 2.3 e l'andamento angolare è rappresentato nei grafici 2.5, 2.6, in cui vengono sovrapposti i risultati ottenuti dai dati sperimentali e dalla simulazione MC.

Tabella 2.3: Risultati efficienza - Metodo C (taglio LOOSE)

| $\Delta\theta(^{\circ})$ | LOOSE           |                   |                 |                   |
|--------------------------|-----------------|-------------------|-----------------|-------------------|
|                          | Eff. Id. $K^+$  |                   | Eff. Id. $K^-$  |                   |
|                          | Dati            | MC                | Dati            | MC                |
| 30-40                    | $0.85 \pm 0.04$ | $0.81 \pm 0.03$   | $0.83 \pm 0.04$ | $0.745 \pm 0.019$ |
| 40-50                    | $0.87 \pm 0.04$ | $0.852 \pm 0.017$ | $0.84 \pm 0.03$ | $0.771 \pm 0.016$ |
| 50-60                    | $0.86 \pm 0.03$ | $0.86 \pm 0.02$   | $0.87 \pm 0.03$ | $0.795 \pm 0.015$ |
| 60-70                    | $0.84 \pm 0.03$ | $0.865 \pm 0.019$ | $0.88 \pm 0.03$ | $0.803 \pm 0.014$ |
| 70-80                    | $0.85 \pm 0.03$ | $0.86 \pm 0.02$   | $0.86 \pm 0.03$ | $0.786 \pm 0.019$ |
| 80-90                    | $0.84 \pm 0.04$ | $0.86 \pm 0.02$   | $0.81 \pm 0.04$ | $0.794 \pm 0.016$ |
| 90-100                   | $0.80 \pm 0.04$ | $0.85 \pm 0.03$   | $0.80 \pm 0.04$ | $0.757 \pm 0.017$ |
| 100-110                  | $0.77 \pm 0.05$ | $0.81 \pm 0.02$   | $0.76 \pm 0.04$ | $0.727 \pm 0.019$ |
| 110-120                  | $0.70 \pm 0.05$ | $0.72 \pm 0.03$   | $0.70 \pm 0.04$ | $0.66 \pm 0.02$   |

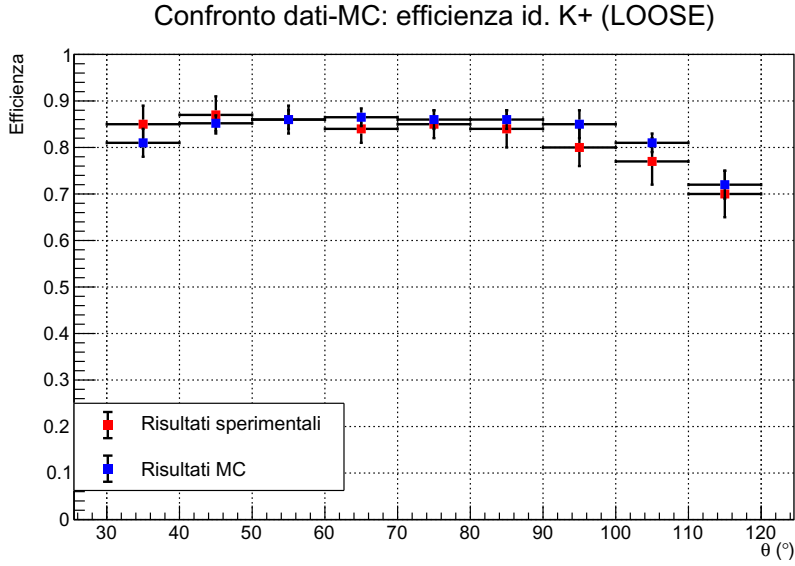


Figura 2.5: Andamento angolare efficienza identificazione  $K^+$  con taglio LOOSE. Confronto dati sperimentali - MC.

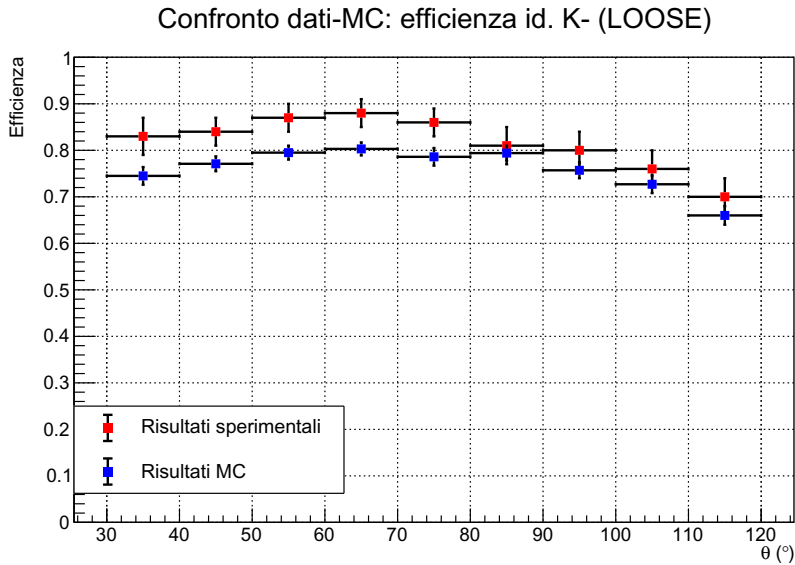


Figura 2.6: Andamento angolare efficienza identificazione  $K^-$  con taglio LOOSE. Confronto dati sperimentali - MC.

Si riscontra un ottimo accordo tra risultati sperimentali e MC: in gran parte degli intervalli la compatibilità tra i due valori è entro  $\sigma$ , e comunque in tutti è entro  $2\sigma$ .

In tutti i campioni si osserva tuttavia un andamento decrescente dell'efficienza nella regione di produzione in avanti ( $\theta > 90^\circ$ ). Tale effetto non era stato evidenziato in precedenti studi dell'efficienza del TOP [4]. Una parziale spiegazione può essere data dal fatto che in questa analisi, a differenza degli studi precedenti, il range angolare in cui opera il TOP è stato diviso molto finemente, considerando nove intervalli di ampiezza  $\Delta\theta = 10^\circ$ . Eventuali variazioni locali dell'efficienza appaiono quindi in modo evidente, cosa che non accade se si considerano intervalli angolari più ampi, in quanto gli effetti risultano mediati. Inoltre, il numero di eventi che risultano costituire il segnale ( $\#C_i$ ) dipende fortemente dal tipo di taglio che viene effettuato sulla PID, come verrà discusso in seguito. In questa analisi è stata valutata l'efficienza relativa, mentre nei precedenti studi era stata valutata quella assoluta (intesa come precedentemente indicato): è possibile che le due presentino un andamento non sovrapponibile.

In ogni caso l'andamento decrescente, seppur chiaramente presente, è confinato entro valori di efficienza che si discostano al più del 10% dall'efficienza media.

### 2.3.2 ALL

Nel caso dei campioni ALL il polinomio di primo grado non è sufficiente per interpolare adeguatamente il fondo, quindi si utilizza un polinomio di secondo grado, con il quale i valori risultanti di  $\chi^2$  consentono di accettare i risultati del fit a C.L. fissato.

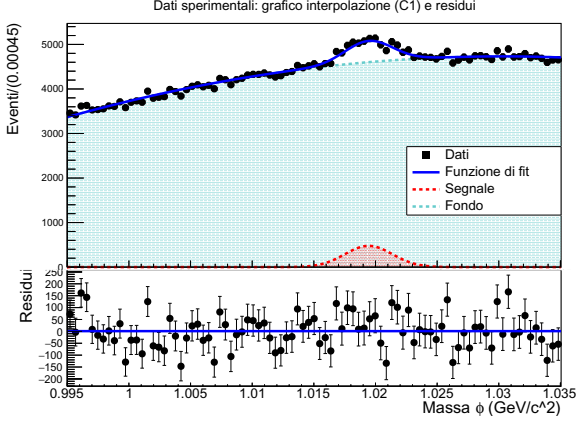


Figura 2.7: Sopra: picco di massa invariante della  $\phi$  nell'intervallo  $60^\circ$ -  $70^\circ$ , dati sperimentali, taglio ALL, campione C1, vincolo su  $\theta_{K^+}$ . In blu il fit totale, in rosso il segnale, in azzurro il fondo. Sotto: residui.

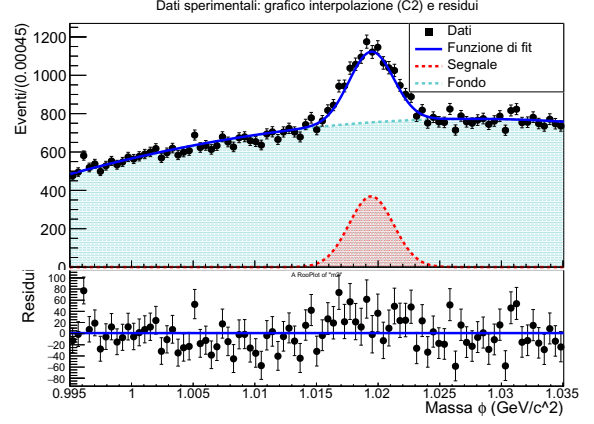


Figura 2.8: Sopra: picco di massa invariante della  $\phi$  nell'intervallo  $60^\circ$ -  $70^\circ$ , dati sperimentali, taglio ALL, campione C2, vincolo su  $\theta_{K^+}$ . In blu il fit totale, in rosso il segnale, in azzurro il fondo. Sotto: residui.

Anche nel caso del taglio ALL la funzione utilizzata per il fit descrive accuratamente la distribuzione dei dati, come si può osservare qualitativamente dai grafici esposti (fig. 2.7, 2.8). L'andamento angolare delle efficienze è mostrato in figura 2.9, 2.10, in cui sono sovrapposti i risultati per i dati sperimentali e per quelli generati dalla simulazione MC.

Tabella 2.4: Risultati efficienza - Metodo C (dati ALL)

| $\Delta\theta(^{\circ})$ | ALL             |                 |                 |                 |
|--------------------------|-----------------|-----------------|-----------------|-----------------|
|                          | Eff. Id. $K^+$  |                 | Eff. Id. $K^-$  |                 |
|                          | Dati            | MC              | Dati            | MC              |
| 30-40                    | $0.71 \pm 0.07$ | $0.66 \pm 0.04$ | $0.55 \pm 0.06$ | $0.7 \pm 0.1$   |
| 40-50                    | $0.82 \pm 0.09$ | $0.70 \pm 0.11$ | $0.81 \pm 0.08$ | $0.84 \pm 0.07$ |
| 50-60                    | $0.77 \pm 0.07$ | $0.84 \pm 0.06$ | $0.86 \pm 0.08$ | $0.79 \pm 0.07$ |
| 60-70                    | $0.70 \pm 0.06$ | $0.93 \pm 0.08$ | $0.92 \pm 0.08$ | $0.92 \pm 0.06$ |
| 70-80                    | $0.71 \pm 0.08$ | $0.79 \pm 0.11$ | $0.75 \pm 0.08$ | $0.74 \pm 0.07$ |
| 80-90                    | $0.86 \pm 0.04$ | $0.8 \pm 0.1$   | $0.80 \pm 0.06$ | $0.80 \pm 0.08$ |
| 90-100                   | $0.85 \pm 0.08$ | $0.82 \pm 0.06$ | $0.78 \pm 0.09$ | $0.88 \pm 0.10$ |
| 100-110                  | $0.77 \pm 0.09$ | $0.65 \pm 0.06$ | $0.86 \pm 0.09$ | $0.86 \pm 0.09$ |
| 110-120                  | $0.59 \pm 0.04$ | $0.54 \pm 0.14$ | $0.54 \pm 0.06$ | $0.72 \pm 0.15$ |



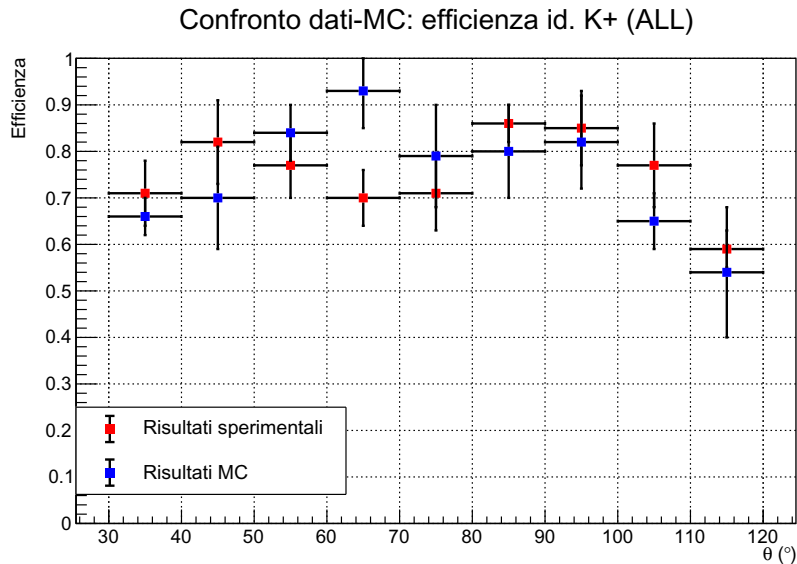


Figura 2.9: Andamento angolare efficienza identificazione  $K^+$  con taglio ALL. Confronto dati sperimentali - MC.

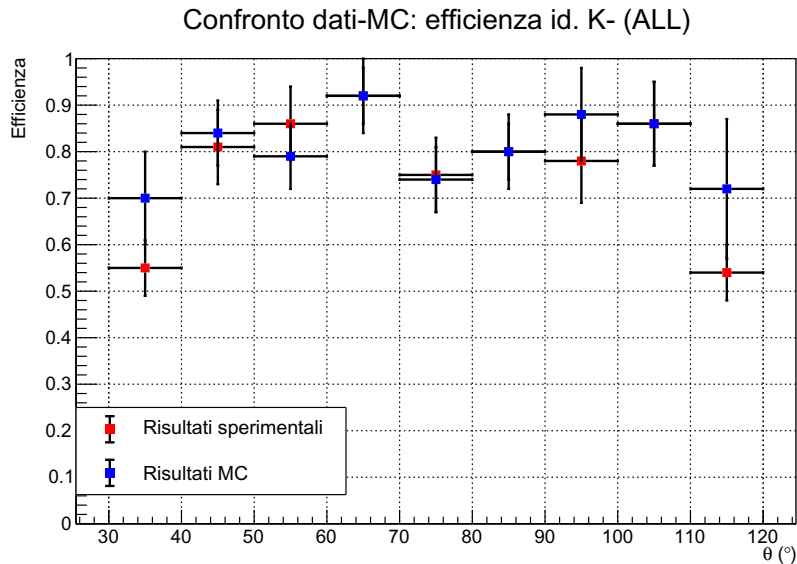


Figura 2.10: Andamento angolare efficienza identificazione  $K^-$  con taglio ALL. Confronto dati sperimentali - MC.

Anche nel caso dei dati ALL si riscontra un buon accordo tra risultati ottenuti dai dati sperimentali e dai MC: in tutti gli intervalli le efficienze risultano compatibili entro  $2\sigma$ . Si sottolinea tuttavia che le barre d'errore associate ai risultati sono sensibilmente maggiori che per i LOOSE.

Le barre d'errore sulle efficienze discendono direttamente dalle barre d'errore sul parametro  $nsig$  determinato dal fit. Nei campioni ALL, come risulta evidente dall'osservazione dei grafici esposti a titolo di esempio (fig. 2.7, 2.8), il fondo è assolutamente preponderante rispetto al segnale (in  $C1$  la frazione di fondo supera il 90%). Il modello utilizzato per descrivere il fondo (polinomio di Chebyshev di secondo grado) è stato assunto sulla base dell'osservazione della distribuzione dei dati e la sua validità risulta confermata dal test del  $\chi^2$  sul fit; non ci sono tuttavia ragioni teoriche che impongano la scelta di tale modello, dunque non stupisce che di fatto non tutti i dati siano in perfetto accordo con esso. Gli scarti dei dati dalla funzione di fit sono alla base dell'errore sui parametri: risulta quindi coerente che nonostante l'elevata statistica a disposizione gli errori sulle efficienze siano piuttosto elevati. Poiché non si riscontrano andamenti sistematici nella distribuzione dei residui, si ritiene che i parametri forniti dal fit siano consistenti e ben descrivano i dati: la barra d'errore risulta quindi rendere conto dell'effettiva incertezza casuale sui valori determinati.

Si osserva inoltre che negli intervalli  $I_1 = [30^\circ - 40^\circ]$  e  $I_9 = [110^\circ - 120^\circ]$  le efficienze risultano più basse che nel resto del range in cui opera il rivelatore ( $I^* = [40^\circ - 110^\circ]$ ). Questa tendenza è presente sia nei dati sperimentali sia nei MC. È possibile che l'efficienza risenta di effetti di bordo dovuti alla presenza dello specchio ( $I_1$ ) e del prisma ( $I_9$ ).

Anche nei due intervalli in esame ( $I_1, I_9$ ) le efficienze calcolate da dati sperimentali e MC risultano compatibili, ma a causa dell'elevata barra d'errore associata ai dati non è possibile dire se i risultati siano realmente sovrapponibili, ovvero se nella simulazione specchio e prisma vengano effettivamente modellizzati nel modo più adeguato.

## 2.4 Osservazioni sulla scelta del metodo

Complessivamente, si ritiene che il metodo migliore per stimare l'efficienza del TOP a partire dai dati in esame sia il metodo **C**.

Sebbene i risultati finali ricavati con i tre metodi non siano molto diversi, e dunque la scelta del metodo non influisca sensibilmente sulle stime fornite, si ritiene che il metodo **C** sia più adeguato degli altri in quanto nella sua applicazione non emergono i problemi concettuali o computazionali evidenziati nel corso dell'analisi per **A** e **B**.

Il metodo **A**, infatti, pur essendo concettualmente corretto e coerente con la trattazione teorica dei decadimenti, non risulta adeguato ai fini del calcolo dell'efficienza perché di fatto il fit non riesce a stimare in modo né preciso né accurato il rapporto segnale/rumore.

Il metodo **B**, al contrario, fornisce risultati molto precisi, ma il modello utilizzato per descrivere il segnale (singola gaussiana su fondo lineare o quadratico) appare troppo semplificato, specialmente per la trattazione dei dati ALL.

Il metodo **C** supera entrambi questi problemi. Da un lato, modellizzando il segnale come somma di due gaussiane si riesce a tenere conto in modo più adeguato della distribuzione effettiva dei dati. Dall'altro, non utilizzando come parametro il rapporto segnale/rumore, ma il numero di eventi di segnale e rumore, il fit fornisce risultati più soddisfacenti.

Tra le stime fornite per l'efficienza del TOP si considerano quindi valide quelle esposte in tabella 2.3 e 2.4, ricavate con il metodo di analisi **C**.

## Capitolo 3

# Efficienza relativa complessiva

Come precedentemente osservato, nell'intervallo angolare  $I^* = \theta \in [40^\circ - 110^\circ]$  l'efficienza del TOP risulta approssimativamente costante, pertanto in tale range è possibile fornire un'unica stima che rappresenti l'efficienza complessiva del rivelatore.

Per calcolare questa efficienza complessiva si applica il metodo **C**, che è risultato il più efficace, ai dati ALL considerando un unico campione comprendente tutti i dati aventi  $\theta \in [40^\circ - 110^\circ]$ .

Si espongono di seguito i grafici dei fit effettuati ai fini del calcolo dell'efficienza nell'identificazione dei  $K^-$ ; un'analisi equivalente è stata fatta per i  $K^+$ .

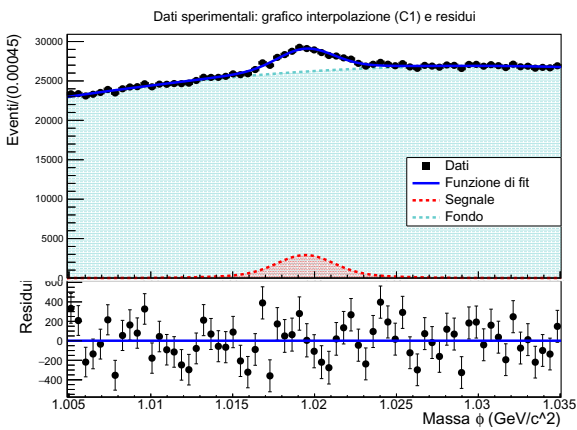


Figura 3.1: Sopra: picco di massa invariante della  $\phi$ , dati sperimentali, campione C1, vincolo su  $\theta_{K^-}$ . In blu il fit totale, in rosso il segnale, in azzurro il fondo. Sotto: residui.

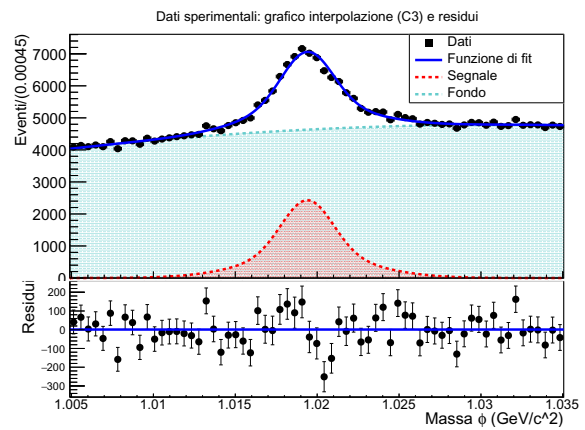


Figura 3.2: Sopra: picco di massa invariante della  $\phi$ , dati sperimentali, campione C3, vincolo su  $\theta_{K^-}$ . In blu il fit totale, in rosso il segnale, in azzurro il fondo. Sotto: residui.

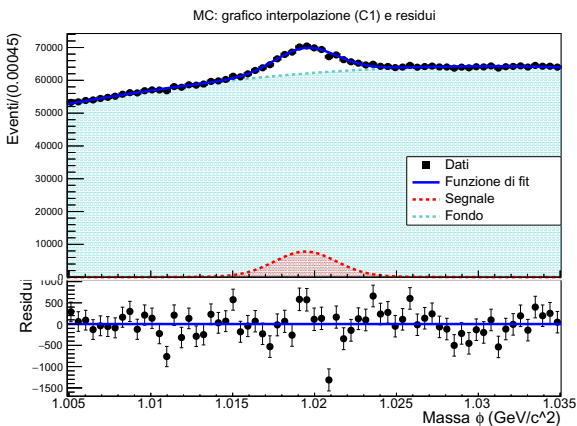


Figura 3.3: Sopra: picco di massa invariante della  $\phi$ , MC, campione C1, vincolo su  $\theta_{K^-}$ . In blu il fit totale, in rosso il segnale, in azzurro il fondo. Sotto: residui.

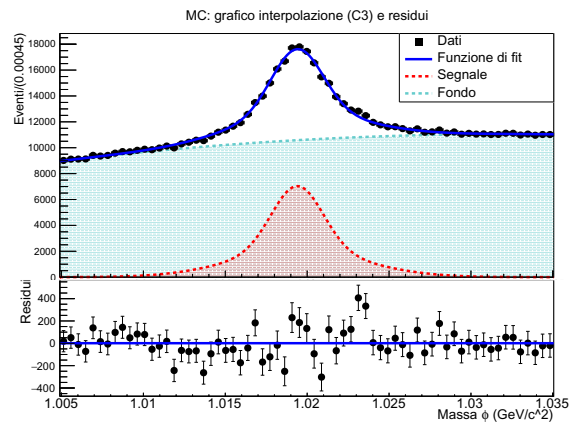


Figura 3.4: Sopra: picco di massa invariante della  $\phi$ , MC, campione C3, vincolo su  $\theta_{K^-}$ . In blu il fit totale, in rosso il segnale, in azzurro il fondo. Sotto: residui.

Un approccio alternativo per fornire la medesima stima consiste nell'effettuare una media pesata tra le efficienze determinate nei singoli intervalli angolari ( $\Delta\theta_i$ ), precedentemente calcolate ed esposte in tabella 2.4. Si sottolinea che è possibile procedere anche in questo modo poiché, per tutti i campioni, nel range  $I^*$  le efficienze sono risultate compatibili tra loro.

L'efficienza complessiva del rivelatore nell'identificazione di  $K^+$  e  $K^-$  risulta rispettivamente:

Tabella 3.1: Risultati efficienza in  $I^*$

| Metodo                                      | Def.             | ALL             |                 |                 |                 |
|---|------------------|-----------------|-----------------|-----------------|-----------------|
|   |                  | Eff. Id. $K^+$  |                 | Eff. Id. $K^-$  |                 |
|   |                  | Dati            | MC              | Dati            | MC              |
| Unico fit in $I^* = [40^\circ - 110^\circ]$ | $\epsilon_{I^*}$ | $0.79 \pm 0.08$ | $0.82 \pm 0.05$ | $0.85 \pm 0.09$ | $0.86 \pm 0.08$ |
| Media pesata $\epsilon_{\Delta\theta^i}$    | $\bar{\epsilon}$ | $0.79 \pm 0.03$ | $0.79 \pm 0.03$ | $0.82 \pm 0.03$ | $0.83 \pm 0.03$ |

I risultati ottenuti con i due metodi risultano sempre compatibili entro  $\sigma$ .

Si osserva che l'errore associato ai risultati ottenuti effettuando la media pesata tra le efficienze nei singoli intervalli angolari ( $\bar{\epsilon}$ ) è sensibilmente minore di quello ricavato dall'unico fit completo su  $I^*$  ( $\epsilon_{I^*}$ ). Ciò da un lato non stupisce, in quanto la media pesata per costruzione riduce significativamente l'errore, rispetto a quello associato ai valori mediati ( $\epsilon_{\Delta\theta_i}$ , tab. 2.4). Dall'altro, tuttavia, bisogna considerare che effettuando un unico fit in tutto  $I^*$  si ha a disposizione una statistica molto più elevata che in ciascun  $\Delta\theta_i$ . Ci si aspetterebbe quindi che l'errore associato alle  $\epsilon_{I^*}$  fosse minore di quelli associati alle efficienze nei singoli intervalli angolari (tab. 2.4). Si osserva che, invece, risultano confrontabili: l'errore non cala significativamente aumentando la statistica. Ciò significa che l'errore associato ai risultati  $\epsilon_{I^*}$ ,  $\epsilon_{\Delta\theta_i}$  rappresenta la loro reale incertezza, derivante dall'effettiva dispersione dei dati, la quale mantiene lo stesso andamento anche all'aumentare della statistica.

Si ritiene quindi che la media pesata sottostimi l'errore associato all'efficienza nell'intervallo  $I^*$ .

Si considerano allora come risultati finali per l'efficienza del TOP in  $I^* = \theta \in [40^\circ - 110^\circ]$  quelli ottenuti effettuando un unico fit nell'intervallo ( $\epsilon_{I^*}$  in tab. 3.1).

## Capitolo 4

# Efficienza assoluta in funzione delle selezioni sulla PID

Nell'analisi svolta nei capitoli precedenti è stata considerata l'efficienza relativa del TOP nell'identificazione dei Kaoni rispetto ai Pioni. Le variabili di PID prese in esame finora, sulle quali sono stati imposti i vincoli necessari ai fini dell'analisi, sono PairK e PairPi, definite come indicato nella sezione 1.3, che restituiscono la probabilità di un'ipotesi di particella ( $K$  o  $\pi$ ) rispetto all'altra. PairK e PairPi, cioè, valutano solamente due ipotesi per ciascuna particella in esame, mutuamente esclusive (che sia un Kaone o un Pione).

Le efficienze sono state quindi calcolate come

$$\epsilon_{K^+} = \frac{\# C2 (K^- \text{PairK} > K^- \text{PairPi})}{\# C1 (\text{No vincoli su PID})} \quad \epsilon_{K^-} = \frac{\# C3 (K^+ \text{PairK} > K^+ \text{PairPi})}{\# C1 (\text{No vincoli su PID})}$$

È tuttavia possibile dare una definizione operativa differente dell'efficienza del rivelatore.

In particolare, al posto che l'efficienza relativa, ottenuta imponendo  $K^\pm \text{PairK} > K^\pm \text{PairPi}$ , è possibile valutare l'efficienza assoluta imponendo che la probabilità che la particella in esame sia un Kaone sia maggiore di una certa soglia fissata. Inoltre, come variabile di PID su cui imporre il vincolo è possibile utilizzare la PIDE (Particle IDentification for Expert), che non si limita a considerare le due ipotesi Kaone/Pione per la particella in esame, ma al contrario considera come alternative possibili le ipotesi Kaone, Pione, Elettrone, Protone, Muone, Deuterio. In particolare, la variabile PIDEk, definita come

$$\text{PIDEk} = \frac{\mathcal{L}_K}{\mathcal{L}_K + \mathcal{L}_\pi + \mathcal{L}_e + \mathcal{L}_p + \mathcal{L}_\mu + \mathcal{L}_d}$$

con  $\mathcal{L}$  massima verosimiglianza, restituisce la probabilità che la particella sia un Kaone e non una delle altre alternative possibili, sulla base delle informazioni fornite dal solo rivelatore TOP.

Per calcolare l'efficienza assoluta del rivelatore è quindi necessario imporre  $\text{PIDEk} > \lambda$ , con  $\lambda \in [0, 1]$  soglia fissata.

Vengono quindi ridefiniti i campioni:

- **C1**: nessun vincolo sulla PID
- **C2**:  $K^- \text{PIDEk} > \lambda$  e nessun vincolo su  $K^+ \text{ID}$
- **C3**:  $K^+ \text{PIDEk} > \lambda$  e nessun vincolo su  $K^- \text{ID}$

L'efficienza (assoluta) del TOP nell'identificazione di  $K^+$  e  $K^-$  è quindi:

$$\epsilon_{K^+}^\lambda = \frac{\# C2}{\# C1} \quad \epsilon_{K^-}^\lambda = \frac{\# C3}{\# C1}$$

Vengono di seguito esposti i risultati per le efficienze al variare della soglia  $\lambda$  (tabella 4.1). Le stime sono state ricavate applicando il metodo **C** direttamente ai campioni di dati ALL.

Tabella 4.1: Risultati efficienza al variare di  $\lambda$

| $\lambda$ | ALL             |                 |                 |                 |
|-----------|-----------------|-----------------|-----------------|-----------------|
|           | Eff. Id. $K^+$  |                 | Eff. Id. $K^-$  |                 |
|           | Dati            | MC              | Dati            | MC              |
| 0.5       | $0.61 \pm 0.04$ | $0.61 \pm 0.03$ | $0.65 \pm 0.05$ | $0.60 \pm 0.03$ |
| 0.6       | $0.60 \pm 0.04$ | $0.60 \pm 0.03$ | $0.64 \pm 0.05$ | $0.59 \pm 0.04$ |
| 0.7       | $0.59 \pm 0.04$ | $0.59 \pm 0.03$ | $0.62 \pm 0.05$ | $0.59 \pm 0.02$ |
| 0.8       | $0.58 \pm 0.04$ | $0.58 \pm 0.03$ | $0.61 \pm 0.05$ | $0.57 \pm 0.02$ |
| 0.9       | $0.56 \pm 0.04$ | $0.57 \pm 0.03$ | $0.60 \pm 0.08$ | $0.53 \pm 0.02$ |
| 0.95      | $0.55 \pm 0.03$ | $0.55 \pm 0.04$ | $0.59 \pm 0.04$ | $0.52 \pm 0.02$ |

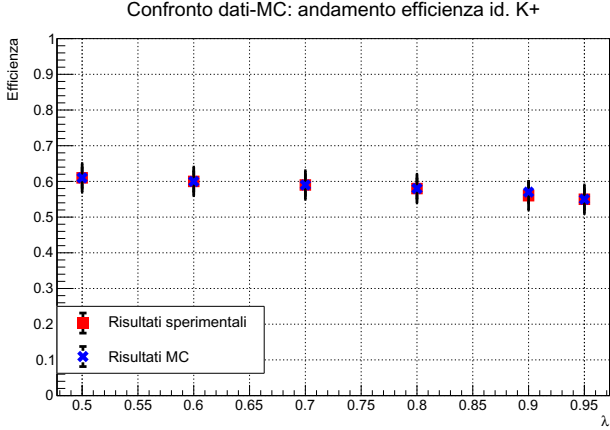


Figura 4.1: Andamento efficienza identificazione  $K^+$  al variare di  $\lambda$ . Confronto dati sperimentali - MC.

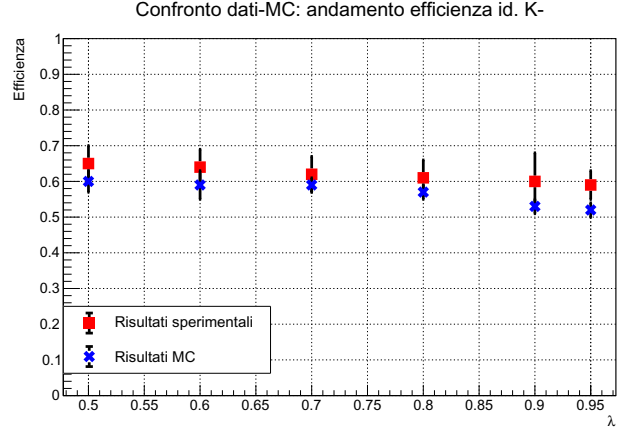


Figura 4.2: Andamento efficienza identificazione  $K^-$  al variare di  $\lambda$ . Confronto dati sperimentali - MC.

Per le efficienze nell'identificazione dei  $K^+$  la compatibilità tra risultati sperimentali e MC è ottima, i valori sono perfettamente sovrapponibili. Nel caso dei  $K^-$ , si osserva una maggiore discrepanza tra dati e MC; nonostante ciò i valori risultano compatibili entro  $\sigma$  per tutti i  $\lambda$ , eccetto nel caso di  $\lambda = 0.95$ , in cui la compatibilità sale a  $c \approx 1.6$  (comunque buona).

In fase di analisi si è osservato che i fit effettuati con questo vincolo sulle variabili di PID convergono in modo molto più stabile rispetto a quelli effettuati nelle sezioni precedenti. In particolare, si è riscontrata una sensibilità ai parametri iniziali nettamente minore. Minori sono anche le barre d'errore associate ai parametri forniti dal fit, e conseguentemente quelle sulle stime dell'efficienza, che risultano ridotte di un fattore 2 rispetto a quelle precedentemente ricavate ( $\frac{\sigma_\epsilon}{\epsilon}$  passa da  $\approx 10\%$  a  $\approx 5\%$ ).

Si osserva che nel complesso le efficienze assolute stimate con questo metodo risultano più basse di quelle relative precedentemente calcolate. Tale risultato non stupisce: utilizzando la PIDEk si introducono ipotesi di particella che la PairK non contemplava, pertanto per la particella aumentano le alternative possibili all'ipotesi *Kaone*, e con esse la probabilità di non rispettare la condizione imposta.

La definizione operativa di  $\epsilon_{K^\pm}^\lambda$  risulta più fedele, rispetto a quella che fa uso di PairK, alla definizione classica di efficienza, secondo cui  $\epsilon_K = \frac{\# K \text{ rivelate}}{\# K \text{ incidenti}}$ . Nella definizione classica, infatti, non si pongono vincoli sulle ipotesi di particella alternative rispetto a quella in esame. Una definizione operativa, quindi, è tanto più simile a quella classica quanto meno stringenti sono le ipotesi alternative.

# Conclusioni

In questo studio è stata valutata l'efficienza del rivelatore TOP: ne è stato esaminato l'andamento in funzione dell'angolo di emissione dei  $K$  e della definizione operativa utilizzata per il calcolo.

L'analisi dei risultati ha permesso di stabilire che il modello più adatto alla descrizione del picco di massa invariante della  $\phi$  è quello che utilizza la somma di due gaussiane di diversa larghezza (segnale), sommate a un polinomio di Chebyshev di primo o secondo grado (fondo).

## Efficienza relativa in funzione dell'angolo

Si è riscontrato un buon accordo (compatibilità entro  $2\sigma$ ) tra i risultati ottenuti dai dati sperimentali e i corrispondenti risultati ottenuti dai dati generati con la simulazione MC. In entrambi i campioni l'efficienza risulta mediamente compresa tra 0.8 e 0.85. In fase di analisi si è tuttavia evidenziata un'estrema sensibilità dei fit, e conseguentemente dei risultati per le efficienze, ai valori iniziali forniti ai parametri. Tale incertezza è rispecchiata dall'ampia barra d'errore associata ai risultati: mediamente  $\frac{\sigma_\epsilon}{\epsilon} \approx 10\%$ .

## Efficienza assoluta in funzione delle selezioni sulla PID

Si è riscontrato un ottimo accordo tra risultati sperimentali e MC (compatibilità entro  $\sigma$  nella quasi totalità dei casi). La convergenza dei fit è risultata molto più stabile che nel caso precedente; parallelamente, le barre d'errore associate ai risultati sono sensibilmente minori ( $\frac{\sigma_\epsilon}{\epsilon}$  tra 4% e 7%). Come previsto, le efficienze stimate con questo metodo ( $\text{PID}_{\text{EK}} > \lambda$ ) risultano inferiori di quelle stimate imponendo  $K^\pm\text{PairK} > K^\pm\text{PairPi}$ : mediamente sono infatti comprese tra 0.55 e 0.60.

Si sottolinea che le stime fornite per l'efficienza del TOP sono estremamente conservative. Tutta l'analisi è stata svolta cercando di non introdurre bias: particolare attenzione è stata posta a non utilizzare variabili di PID che contenessero informazioni derivanti da altri rivelatori.

Questo approccio consente di studiare l'efficienza del TOP come rivelatore singolo, e di fornirne una stima non falsata dalle conoscenze collaterali a disposizione. Inevitabilmente però, ne sottostima l'efficienza all'interno dell'esperimento, contesto in cui di fatto tali conoscenze sono disponibili.

Per caratterizzare in modo completo il rivelatore, quindi, accanto a studi come quello qui presentato, è opportuno valutare l'efficienza complessiva che riesce a raggiungere lavorando in concerto con gli altri rivelatori dell'esperimento.

# Bibliografia

- [1] <https://geant4.web.cern.ch/>.
- [2] <https://root.cern/>.
- [3] <https://root.cern/manual/roofit/>.
- [4] A. Schwartz A. Sangal, S. Sandilya. *Kaon Identification in BelleII using  $\phi \rightarrow K^+K^-$* . March 2019. University of Cincinnati, Cincinnati, Ohio. BELLE2-NOTE-PH-2018-037.
- [5] P. Branchini. *The Belle II Experiment: Status and Prospects*. Universe, 4:101, October 2018.
- [6] C. Cecchi. *The Belle II Experiment: Status and Prospects*. IL NUOVO CIMENTO, 100(249):42, March 2019. BELLE2-TALK-CONF-2019-014.
- [7] M. Paterno and Fermilab. *Calculating efficiencies and their uncertainties*. December 2004. United States. FERMILAB-TM-2286-CD.
- [8] U. Tamponi. *The TOP counter of Belle II: Status and first results*. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 952, May 2019.
- [9] M. Tanabashi *et al.* (Particle Data Group). Review of particle physics. Physical Review D, 98(3), August 2018.