



**UNIVERSITÀ DEGLI STUDI
DI CAGLIARI**

**Facoltà di Scienze Matematiche, Fisiche e Naturali
Corso di Laurea Specialistica in Fisica**

**Applicazione di metodi
multidimensionali
alla selezione di eventi rari in LHCb**

Relatore

Prof. Biagio Saitta

Correlatore

Dott. Walter Bonivento

Tesi di laurea di:

Francesco Dettori

**Anno Accademico
2006-2007**

Introduzione

Uno dei campi più promettenti per lo studio di fisica oltre il Modello Standard (SM) delle particelle elementari sono i decadimenti rari dei mesoni B .

Essi comprendono processi che sono fortemente soppressi o proibiti all'interno del Modello Standard. Per la maggior parte si tratta di decadimenti Cabibbo-soppressi o *Flavour Changing Neutral Currents* (FCNC). La rarità di questi processi fa sì che i contributi dovuti alla presenza di fisica oltre il Modello Standard possano essere dello stesso ordine o maggiori rispetto ai contributi dominanti nel SM. Questo ne fa canali privilegiati per la ricerca di nuova fisica.

Tra i processi proibiti nel SM, vi sono i decadimenti che violano la conservazione del numero leptonico (LV) o del numero leptonico di famiglia (LFV). Questi sono permessi in altri modelli fisici e possono portare quindi a scoperte dirette di nuova fisica.

L'esperimento LHCb, in corso di installazione presso l'acceleratore di particelle LHC a Ginevra, avrà come obiettivo la fisica dei B sia per uno studio della violazione della simmetria CP sia per quanto riguarda i decadimenti rari.

Un aspetto importante è dunque la selezione degli eventi rari interessanti, che dev'essere il più efficace e libera da fondo possibile. Il metodo tipico per effettuare delle selezioni è costituito da una serie di tagli applicati su variabili cinematiche e dinamiche di interesse. Questo metodo tuttavia non sfrutta l'intera informazione contenuta in un evento. Vari metodi multidimensionali possono essere usati per lo sfruttamento di questa informazione e la costruzione di analisi più efficienti.

Il lavoro di questa tesi si è quindi concentrato sullo studio di metodi

multidimensionali tramite i quali effettuare selezioni di eventi. In particolare se ne è considerata l'applicazione allo studio dei decadimenti rari dei mesoni B nell'ambito dell'esperimento LHCb.

Sono stati considerati i decadimenti LFV $B_{d,s}^0 \rightarrow e^\pm \mu^\mp$ e i decadimenti $B_{d,s}^0 \rightarrow \mu^+ \mu^- \gamma$ (FCNC).

Questa tesi è strutturata come segue: nel Capitolo 1 verranno discussi brevemente i modelli teorici alla base dei decadimenti studiati. Di seguito nel Capitolo 2 si illustrerà l'esperimento LHCb, nell'ambito del quale si inserisce questo lavoro. Quindi si mostreranno i metodi multivariati che sono stati studiati, come funzionano e come sono stati implementati (Capitolo 3). L'applicazione di questi metodi ai decadimenti di interesse sarà discussa per $B_{d,s}^0 \rightarrow e^\pm \mu^\mp$ nel Capitolo 4 e per $B_{d,s}^0 \rightarrow \mu^+ \mu^- \gamma$ nel Capitolo 5, traendo le conclusioni di questo studio di seguito nel Capitolo 6.

Indice

Introduzione	i
1 Decadimenti rari del B	1
1.1 Decadimenti con violazione del numero leptonico	1
1.1.1 Il modello di Pati-Salam	2
1.1.2 I decadimenti $B_{d,s}^0 \rightarrow e^\pm \mu^\mp$	5
1.2 Flavour changing neutral currents	6
1.2.1 Il Decadimento $B_s^0 \rightarrow \mu^+ \mu^-$	7
1.2.2 Il Decadimento $B_s^0 \rightarrow \mu^+ \mu^- \gamma$	8
2 L'esperimento LHCb	13
2.1 LHC - Large Hadron Collider	13
2.2 Collisioni protone-protone	14
2.2.1 Produzione di b	15
2.3 Il rivelatore LHCb	16
2.3.1 Il VELO	19
2.3.2 Il RICH	20
2.3.3 Il Trigger Tracker	21
2.3.4 Il magnete	21
2.3.5 Il Tubo del fascio	22
2.3.6 Il sistema di tracciamento	22
2.3.7 I Calorimetri	23
2.3.8 Il sistema per muoni	24
2.3.9 Il trigger	24
2.3.10 Il sistema di <i>computing</i> di LHCb	25

3	Metodi multidimensionali di selezione degli eventi	27
3.1	Metodi multidimensionali	28
3.1.1	Likelihood	28
3.1.2	Decorrelazione delle variabili	31
3.1.3	Approssimazione di proiezione e correlazione (PCA)	32
3.1.4	Fisher	34
3.1.5	Reti neurali artificiali	35
3.1.6	Decision Trees	36
3.2	Training, Ottimizzazione e criterio di Punzi	37
3.2.1	Training	37
3.2.2	Overtraining	39
3.2.3	Ottimizzazione e Criterio di Punzi	40
3.3	Implementazione e calcolo	41
3.3.1	Implementazione	42
3.3.2	Valutazione dei metodi	43
3.3.3	ROOT	44
3.3.4	TMVA	44
4	Studio dei decadimenti $B_{d,s}^0 \rightarrow e^\pm \mu^\mp$	47
4.1	Segnale e Fondi	47
4.1.1	Simulazione degli eventi	48
4.2	Preselezione	49
4.3	Selezione degli eventi	49
4.3.1	Applicazione dei metodi multidimensionali	54
4.4	Strategia di applicazione dell'analisi ai dati reali	56
4.5	Risultati	59
5	Studio del decadimento $B_s^0 \rightarrow \mu^+ \mu^- \gamma$	61
5.1	Segnale e fondo	61
5.1.1	Simulazione degli eventi	62
5.2	Preselezione	63
5.3	Selezione degli eventi	67
5.3.1	Applicazione dei metodi multidimensionali	67

5.4 Risultati	72
6 Conclusioni	73
Bibliografia	75
Ringraziamenti	79

Elenco delle tabelle

2.1	Sezioni d'urto per lo scattering p-p	15
2.2	Frazioni di adroni b ottenute in decadimenti $Z^0 \rightarrow b\bar{b}$ e collisioni $p\bar{p}$ a $\sqrt{s} = 1.8$ TeV.	16
4.1	Statistica degli eventi simulati analizzati	49
4.2	Tagli di preselezione applicati ai dati	49
4.3	Tagli applicati ai dati nella selezione finale	52
4.4	Stima dell' <i>overtraining</i> per i metodi testati.	54
4.5	Tagli applicati alle DLL dei leptoni e livelli di contaminazione relativa del fondo <i>two body</i> rispetto al leptonic	59
5.1	Tagli di preselezione applicati ai dati per la selezione del decadimento $B_s^0 \rightarrow \mu^- \mu^+ \gamma$	65
5.2	Statistica degli eventi e delle combinazioni di segnale e di fondo analizzati	66

Elenco delle figure

1.1	Diagramma di Feynman per il decadimento $B_s^0 \rightarrow e^+\mu^-$ mediato da un <i>leptoquark</i>	5
1.2	Diagrammi di Feynman che contribuiscono al decadimento $B_s^0 \rightarrow \mu^+\mu^-$ all'interno del Modello Standard	9
1.3	Diagrammi di Feynman che contribuiscono al decadimento $B_s^0 \rightarrow \mu^+\mu^-\gamma$ all'interno del Modello Standard	12
2.1	Schema delle strutture sotterranee del Large Hadron Collider .	14
2.2	Diagrammi di Feynman per la produzione di coppie $b\bar{b}$	17
2.3	Il rivelatore LHCb: sezione verticale.	20
3.1	Esempio di distribuzione di Rapporto di Likelihood.	30
3.2	Esempio di distribuzione di rapporto di Likelihood trasformato.	30
3.3	Schema di una rete neurale <i>multilayer</i>	36
4.1	Distribuzioni delle variabili utilizzate nell'analisi (1/2)	52
4.2	Distribuzioni delle variabili utilizzate nell'analisi (2/2)	53
4.3	Distribuzione delle variabili di output per i metodi studiati.	55
4.4	Andamento del numero di eventi di fondo selezionati in 1 fb^{-1} di luminosità integrata rispetto all'efficienza sul segnale (in percentuale), per diversi metodi utilizzati.	56
4.5	Andamento della variabile associata al criterio di Punzi in funzione dell'efficienza sul segnale.	57
4.6	Distribuzione di massa invariante per i diversi fondi del decadimento studiato.	58

5.1	Distribuzione dell'energia trasversa dei fotoni nel decadimento $B_s^0 \rightarrow \mu^+ \mu^- \gamma$	64
5.2	Distribuzione della massa invariante dei due muoni nel decadimento $B_s^0 \rightarrow \mu^+ \mu^- \gamma$	64
5.3	Distribuzioni di alcune delle variabili utilizzate nell'analisi del decadimento $B_s^0 \rightarrow \mu^+ \mu^- \gamma$	68
5.4	Distribuzione delle variabili di output per i metodi studiati. . .	69
5.5	Andamento del numero di eventi di fondo selezionati in un anno di presa dati in funzione dell'efficienza sul segnale. . . .	70
5.6	Andamento del criterio di Punzi in funzione dell'efficienza sul segnale.	71

Capitolo 1

Decadimenti rari del B

Vengono detti rari i decadimenti del B con *branching ratios* dell'ordine di $\sim O(10^{-5})$ o minori, causati da processi o fortemente soppressi nel SM, come le *flavour changing neutral currents* (FCNC), o esplicitamente proibiti come quelli che violano la conservazione del numero leptonico (LFV).

Poiché, visti i *branching ratios*, i contributi di nuova fisica potrebbero essere allo stesso ordine di quelli del Modello Standard (SM), i decadimenti rari dei mesoni B possono essere sfruttati per fornire prove di fisica oltre il SM. Inoltre, qualora le previsioni teoriche del Modello Standard fossero confermate, essi possono essere utili per la misura delle costanti di decadimento adroniche.

In vista dei due canali studiati in questo lavoro, $B_{d,s}^0 \rightarrow e^\pm \mu^\mp$ e $B_s^0 \rightarrow \mu^+ \mu^- \gamma$, di seguito verranno discussi in primo luogo i decadimenti LFV e quindi quelli che coinvolgono FCNC.

1.1 Decadimenti con violazione del numero leptonico

La conservazione del numero leptonico, fondamentale nel Modello Standard, spesso è superata all'interno di altri modelli teorici. Tra questi, alcuni predicono solo la violazione del numero leptonico dei famiglia (L_e, L_μ, L_τ), altri predicono anche la violazione del numero leptonico totale (L).

Di seguito verrà descritto un modello semplice che predice questa viola-

zione nell'ambito di una simmetria maggiore che coinvolge tutti i fermioni elementari.

1.1.1 Il modello di Pati-Salam

Una delle proprietà non spiegate del Modello Standard delle interazioni deboli e forti è il perché alcuni fermioni, i quark, interagiscono in maniera forte mentre altri, i leptoni, no.

L'esperienza insegna a ricercare delle simmetrie anche laddove non sono manifeste, e questo può portare a pensare che ad un qualche livello, o meglio, in qualche regime di energia vi sia una simmetria tra quark e leptoni. Questa simmetria potrebbe quindi essere *rotta* alle basse energie, così come la simmetria tra l'interazione debole e quella elettromagnetica viene meno alla scala di energia di circa 250 GeV.

Se inoltre si pensa che questa possa essere una simmetria di gauge locale, ne deriva un nuovo tipo di interazione tra leptoni e quark, e delle nuove particelle, dette *leptoquark*, che ne sono i mediatori.

Questa è, in breve, l'idea di fondo di una serie di teorie riunite sotto il nome di Teorie della Grande Unificazione (GUT - *Grand Unified Theories*). Fattore comune di queste teorie è il prevedere una simmetria più generale di quelle che attualmente vengono usate per spiegare le interazioni delle particelle elementari; questa subisce una procedura di rottura spontanea della simmetria alla scala GUT tipica del particolare modello. Tutte queste teorie prevedono dei leptoquark che abbiano masse dell'ordine della scala GUT.

Pati e Salam [1, 2, 3] proposero una serie di modelli che descrivono questa unificazione di quark e leptoni: il numero leptonico viene in questi pensato e descritto come come il quarto *colore* dell'interazione forte, accoppiando le tre famiglie di leptoni con i tre *flavour* dei quark.

Tutti i modelli di Pati-Salam hanno come base comune il gruppo di simmetria $SU(4)_c$. Esso si *rompe* ad alte energie dando luogo al sottogruppo $SU(3)_c$, che descrive l'ordinaria interazione forte con gluoni *massless*, e generando dei leptoquark massivi.

Considerando il gruppo di simmetria del Modello Standard $SU(3)_c \times$

$SU(2)_L \times U(1)$, la sua estensione è del tipo $SU(4)_c \times SU(2)_L \times G_R$, dove il pedice L , sta ad indicare la chiralità dell'interazione debole e G_R è una simmetria che si instaura ad alti regimi di energia e che può essere pensata come la *giusta* ipercarica. Nell'analisi di questo modello ci si può restringere, comunque, al sottogruppo $SU(4)_c \times SU(2)_L$; le tre generazioni di fermioni si combinano in rappresentazioni $(\mathbf{4}, \mathbf{2})$ di questo gruppo, del tipo:

$$\begin{pmatrix} u_R & u_G & u_B & \nu \\ d_R & d_G & d_B & \ell \end{pmatrix}_i \quad (1.1)$$

dove il pedice i indica le tre generazioni, u , d , ν e ℓ sono rispettivamente i quark di tipo *up* e *down* e i leptoni di ogni generazione ed R , G e B gli indici di colore.

In generale l'accoppiamento delle particelle non sarà ristretto a omologhe della stessa generazione ma, come nel caso dei quark con la matrice CKM, ci sarà una matrice che prevede una procedura di *mixing* tra le generazioni. Si può generalizzare quindi la matrice CKM a 4 dimensioni; in genere questa matrice è supposta circa diagonale con gli accoppiamenti intergenerazionali molto più piccoli di quelli tra elementi della stessa generazione.

L'associazione delle famiglie leptoniche ai flavour dei quark tuttavia non è banale. In particolare sarebbe un pregiudizio associare i leptoni τ sempre all'ultimo flavour (t e b) dei quark; tutte le combinazioni vanno invece considerate [4].

Ora, le particelle nella rappresentazione in (1.1) in generale non saranno autostati di massa. Poiché la rappresentazione delle tre famiglie è uguale è possibile far sì che il leptone carico di ogni famiglia sia nel suo autostato di massa. Contraendo la notazione (nascondendo il colore) si ha:

$$\begin{pmatrix} u & \nu \\ d & \ell \end{pmatrix}_\ell = \begin{pmatrix} u_e & \nu_e \\ d_e & e \end{pmatrix}, \begin{pmatrix} u_\mu & \nu_\mu \\ d_\mu & \mu \end{pmatrix}, \begin{pmatrix} u_\tau & \nu_\tau \\ d_\tau & \tau \end{pmatrix} \quad (1.2)$$

dove i pedici $\ell = e, \mu, \tau$ indicano le particelle di una famiglia non nel loro autostato di massa. Esse saranno scrivibili in funzione degli autostati di massa come:

$$\nu_\ell = \mathcal{N}_{\ell i} \nu_i, \quad u_\ell = \mathcal{U}_{\ell j} u_j, \quad d_\ell = \mathcal{D}_{\ell k} d_k \quad (1.3)$$

dove \mathcal{N} , \mathcal{U} e \mathcal{D} sono le matrici unitarie del *mixing*.

In questa notazione, la Lagrangiana dell'interazione della corrente di carica debole con i bosoni W si può scrivere:

$$\begin{aligned}\mathcal{L}_W &= \frac{g}{2\sqrt{2}}[(\bar{\nu}_\ell\gamma^\mu(1-\gamma^5)\ell) + (\bar{u}_\ell\gamma^\mu(1-\gamma^5)d_\ell)]W_\mu^* + h.c. = \\ &= \frac{g}{2\sqrt{2}}[\mathcal{N}_{\ell i}^*(\bar{\nu}_i\gamma^\mu(1-\gamma^5)\ell) + \mathcal{U}_{\ell j}^*\mathcal{D}_{\ell k}\bar{u}_j\gamma^\mu(1-\gamma^5)d_k)]W_\mu^* + h.c.\end{aligned}\quad (1.4)$$

dove g è la costante di accoppiamento del gruppo $SU(2)_L$ e $h.c.$ sta per *hermitiano coniugato*. Si può dunque vedere che la matrice CKM standard è $V = \mathcal{U}^\dagger \mathcal{D}$.

Dalla rottura spontanea di simmetria del gruppo $SU(4)_c$ in $SU(3)_c$ alla scala di massa M_{LQ} , derivano sei bosoni vettori massivi che si separano dal 15-pletto dei campi di gauge, e generano tre leptoquark carichi e *colorati*. La loro interazione con i fermioni ha la forma:

$$\mathcal{L}_{LQ} = \frac{g_S(\mu)}{\sqrt{2}}[\mathcal{D}_{\ell j}(\bar{\ell}\gamma^\mu d_j^c) + \mathcal{N}^\dagger \mathcal{U}_{ik}(\bar{\nu}_i\gamma^\mu u_k^c)]X_\mu^c + h.c.\quad (1.5)$$

dove è nuovamente scritto esplicitamente l'indice di colore c . L'accoppiamento $g_S(\mu)$ può essere scritto in funzione della costante di accoppiamento forte, in quanto alla scala della rottura spontanea di simmetria deve avvenire che $g_S^2(M_{LQ})/4\pi = \alpha_S(M_{LQ})$.

Nelle regioni di momento trasferito $q \ll M_{LQ}$ allora questa lagrangiana media delle interazioni a 4 fermioni (2 quark e 2 leptoni, con diagrammi del tipo mostrato in Figura 1.1).

In realtà per poter scrivere la lagrangiana di interazione efficace occorre tenere conto di correzioni di QCD, si può dimostrare tuttavia che queste appaiono semplicemente come un fattore in più nei termini di interazione scalare e pseudoscalare, del tipo

$$Q(\mu) = \left(\frac{\alpha_S(\mu)}{\alpha_S(M_{LQ})}\right)^{4/b}\quad (1.6)$$

dove $b = 11 - \frac{2}{3}n_f$ e n_f è il numero medio di flavour di quark alla scala μ , cioè $n_f = 6$ e $b = 7$ per $M_X^2 \gg m_t^2$ [5].

Una ricerca diretta dei leptoquark, come coppie prodotte nei collisionatori, attualmente non ha portato a nessuna scoperta. Di conseguenza la ricerca

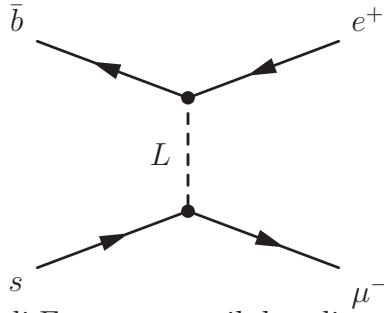


Figura 1.1: Diagramma di Feynman per il decadimento $B_s^0 \rightarrow e^+ \mu^-$ mediato da un *leptoquark*.

si sposta verso reazioni mediate da queste particelle, permesse da questi modelli e altrimenti proibite (o soppresse) nel modello standard: la presenza o meno di questi decadimenti può porre dei limiti stringenti ai parametri di queste teorie, in particolare alle masse dei leptoquark e quindi alla scala di unificazione [6].

In particolare nel modello di Pati-Salam l'interazione con i leptoquark conserva la differenza tra numeri barionico e leptonico ($B - L$) e il numero di fermioni, perciò non può mediare il decadimento di nucleoni. Le transizioni puramente leptoniche del tipo $\mu \rightarrow e\gamma$ o $\mu N \rightarrow eN$ e il *mixing* dei mesoni sono effetti del livello ad un *loop* e sono proporzionali al *mixing* tra le generazioni. Di conseguenza gli effetti più evidenti per la verifica di questi modelli sono i decadimenti, al *tree-level*, dei mesoni neutri. Tramite questi decadimenti si possono porre dei limiti inferiori alle masse dei leptoquark.

1.1.2 I decadimenti $B_{d,s}^0 \rightarrow e^\pm \mu^\mp$

I decadimenti $B_{d,s}^0 \rightarrow e^\pm \mu^\mp$ sono permessi all'interno del modello di Pati-Salam tramite interazioni descritte da diagrammi come quello mostrato in Figura 1.1.

Il *branching ratio* per questi decadimenti è espresso, in funzione della

massa dei leptoquark secondo la seguente formula:

$$\begin{aligned}
BR(B_{d,s}^0 \rightarrow e^\pm \mu^\mp) &= \Gamma(B_{d,s}^0 \rightarrow e^\pm \mu^\mp) \cdot \frac{2\pi\tau_{B(d,s)}}{h} \quad ; \\
\text{con } \Gamma(B_{d,s}^0 \rightarrow e^\pm \mu^\mp) &= \pi\alpha_S^2(M_{LQ}) \frac{F_{B(s,d)}^2 m_{B(s,d)}^3 R^2 (F_{mix}^{d,s})^2}{M_{LQ}^4} \quad (1.7) \\
\text{dove } R &= \frac{m_{B(s,d)}}{m_b} \left(\frac{\alpha_S(M_{LQ})}{\alpha_S(m_t)} \right)^{-4/7} \left(\frac{\alpha_S(m_t)}{\alpha_S(m_b)} \right)^{-12/23}
\end{aligned}$$

e $F_{B(s,d)}^2$, $m_{B(s,d)}$ e $\tau_{B(d,s)}$ sono rispettivamente le costanti di decadimento, le masse e i tempi di vita di $B_{d,s}^0$, m_b e m_t le masse *running* dei quark b e t, e $\alpha_S(\mu)$ la costante di accoppiamento dell'interazione forte alla scala μ . Il fattore $F_{mix}^{d,s}$ tiene conto del *mixing* tra le varie generazioni e, per esempio, per il decadimento $B_s^0 \rightarrow e^\pm \mu^\mp$ vale

$$F_{mix}^s = (\mathcal{D}_{es}\mathcal{D}_{ub}^* + \mathcal{D}_{eb}\mathcal{D}_{us}^*) \quad (1.8)$$

dove \mathcal{D}_{ij} è l'elemento della matrice unitaria che tiene conto del *mixing* dei leptoni e dei quark di tipo *down*.

Attualmente i migliori limiti superiori per i *branching ratios* sono i seguenti: $BR(B_d^0 \rightarrow e^\pm \mu^\mp) < 1.7 \cdot 10^{-7}$ [7] e $BR(B_s^0 \rightarrow e^\pm \mu^\mp) < 6.1 \cdot 10^{-6}$ [8]. La notazione $B_d^0 \rightarrow e^\pm \mu^\mp$ indica uno dei decadimenti $B_d^0 \rightarrow e^+ \mu^-$ e $B_d^0 \rightarrow e^- \mu^+$ e non implica la somma dei due, né la somma dei due col metodo della coniugazione di carica.

Si noti che anche altri modelli teorici predicono questo tipo di decadimento, come ad esempio modelli Seesaw Super-simmetrici mediati da Higgs o modelli con neutrini pesanti; tuttavia questi modelli predicono limiti superiori ai *branching ratios* molto inferiori alla sensibilità dell'esperimento LHCb.

1.2 Flavour changing neutral currents

Per *Flavour Changing Neutral Currents* (FCNC), si intende reazioni tra fermioni di *flavour* differente, ma carica identica. Questo tipo di reazioni (Es. quark $s \rightarrow d$) sono fortemente sopresse all'interno del Modello Standard

per via del Meccanismo di Glashow-Iliopoulos-Maiani (GIM), che fa sì che i termini che contribuirebbero a tali reazioni al livello dominante si cancellino. All'interno del Modello Standard dunque, queste reazioni possono avvenire solo a livelli superiori.

Questo fa sì che le reazioni FCNC siano molto sensibili alla presenza di nuova fisica oltre il Modello Standard; infatti eventuali nuove particelle, pur con accoppiamenti bassi, potrebbero dar luogo per queste reazioni a contributi comparabili, se non superiori a quelli del Modello Standard.

1.2.1 Il Decadimento $B_s^0 \rightarrow \mu^+ \mu^-$

Tra le reazioni FCNC, vi sono i decadimenti leptonic dei mesoni B . Da questi può essere ricavato un importante parametro dei mesoni quale la costante di decadimento leptonica f_B .

I decadimenti leptonic puri, del tipo $B_{s,d}^0 \rightarrow \ell^+ \ell^-$, tuttavia sono soppressi per questioni di elicità. In particolare se i due leptoni fossero privi di massa essi sarebbero completamente proibiti poiché i due leptoni dovrebbero essere emessi con uguale elicità, e quindi chiralità essendo massless, mentre leptoni e antileptoni privi di massa hanno chiralità definita ed opposta. Per questi motivi, se si considera il B_s^0 , i *branching ratios* teorici per questi decadimenti sono:

$$\begin{aligned} BR(B_s^0 \rightarrow e^+ e^-) &\sim 10^{-14} \quad , \\ BR(B_s^0 \rightarrow \mu^+ \mu^-) &\sim 10^{-9} \quad , \\ BR(B_s^0 \rightarrow \tau^+ \tau^-) &\sim 10^{-6} \quad . \end{aligned}$$

Il decadimento in due leptoni τ , pur essendo favorito dalla minore soppressione in elicità, è difficilmente utilizzabile a causa della bassa efficienza di rivelazione.

Il decadimento, tra questi, più interessante, risulta quindi il $B_s^0 \rightarrow \mu^+ \mu^-$. Esso può avvenire tramite i diagrammi di Figura 1.2.

La teoria di questo tipo di transizioni può essere discussa in maniera del tutto generale ricorrendo all'hamiltoniana efficace ottenuta tramite *operator*

product expansion, portando a:

$$\mathcal{H}_{eff}^q = -\frac{4G_F}{\sqrt{2}}V_{tb}V_{ts}^* \sum_{i=1}^{11} C_i(\mu)\mathcal{O}_i^s(\mu) \quad (1.9)$$

dove i \mathcal{O}_i^s sono gli operatori locali rinormalizzati e V_{ij} gli elementi della matrice CKM. I coefficienti di Wilson C_i possono essere calcolati in teoria perturbativa e contenere la fisica rilevante a breve distanza; in particolare essi possono contenere ogni effetto di potenziale nuova fisica [9]. La scala di rinormalizzazione μ può essere vista come la linea di demarcazione tra gli effetti a breve e lunga distanza, e per questo tipo di decadimenti dev'essere scelta dell'ordine di m_b .

Il decadimento $B_s^0 \rightarrow \mu^+\mu^-$ può essere descritto usando tre soli operatori:

$$\begin{aligned} \mathcal{O}_P^s &= (\bar{s}\gamma_5 b)(\bar{\mu}\gamma_5\mu) \\ \mathcal{O}_P^{s'} &= (\bar{s}\gamma_5 b)(\bar{\mu}\mu) \\ \mathcal{O}_A^s &= (\bar{s}\gamma^\alpha\gamma_5 b)(\bar{\mu}\gamma^\alpha\gamma_5\mu) \end{aligned} \quad (1.10)$$

introducendo i coefficienti di Wilson $C_{P,A}^{s,s'}$, si ha il seguente *branching ratio*:

$$BR(B_s^0 \rightarrow \mu^+\mu^-) = \frac{G_F^2}{8\pi}\tau_B f_B^2 m_B^3 \sqrt{1 - \frac{4m_\mu^2}{m_B^2}} \left\{ \left| C_P^s - \frac{m_\mu}{m_B} C_A^s \right| + \left(1 - \frac{4m_\mu^2}{m_B^2} \right) |C_P^{s'}|^2 \right\} \quad (1.11)$$

Questo porta ad un valore del *branching ratio* dell'ordine di $BR(B_s^0 \rightarrow \mu^+\mu^-) = (3.5 \pm 1.0) \cdot 10^{-9}$ [9].

All'interno dell'esperimento LHCb questo decadimento verrà ampiamente studiato, in particolare si può vedere la referenza [10].

1.2.2 Il Decadimento $B_s^0 \rightarrow \mu^+\mu^-\gamma$

Il decadimento $B_s^0 \rightarrow \mu^+\mu^-\gamma$ può avvenire all'interno del Modello Standard come un fotone che viene emesso da una delle tracce cariche della reazione $B_s^0 \rightarrow \mu^+\mu^-$.

L'hamiltoniana efficace che governa questo processo è dunque la medesima di (1.9), con gli operatori già menzionati, alla quale va aggiunto il contributo del fotone.

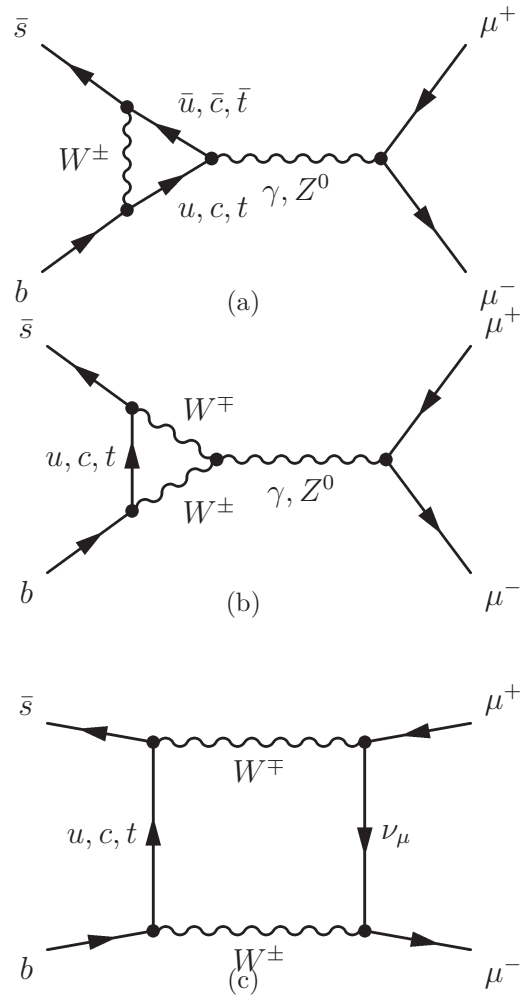


Figura 1.2: Diagrammi di Feynman che contribuiscono al decadimento $B_s^0 \rightarrow \mu^+ \mu^-$ all'interno del Modello Standard

Consideriamo inizialmente i contributi a breve distanza. Non tutte le correzioni sono dello stesso ordine. Si può dimostrare infatti che l'ampiezza del diagramma con un fotone attaccato ad uno dei rami leptonici in uscita, per questioni di elicità, dev'essere proporzionale alla massa del leptone, ed è dunque soppressa.

L'ampiezza con il fotone attaccato ad una delle gambe interne cariche è anch'essa soppressa di un fattore m_b^2/M_W^2 nel coefficiente di Wilson, poiché l'operatore risultante è due ordini di grandezza maggiore (di dimensione 8 anziché 6).

Ne risulta quindi che i contributi dominanti a piccola distanza vengono dai diagrammi con il fotone attaccato ad uno dei due quark dello stato iniziale, alcuni di questi diagrammi son mostrati in Figura 1.3.

Ci sono inoltre contributi a lunga distanza. Per esempio ci sono decadimenti a cascata del tipo $B_s \rightarrow \phi\gamma$ con $\phi \rightarrow \mu^+\mu^-$. La ϕ può inoltre essere sostituita da altre risonanze quali ϕ' , J/ψ etc. Questi contributi, dell'ordine di 10^{-9} in *branching ratio*, appaiono come picchi nella massa invariante dei due muoni, e possono in questo modo essere esclusi nel caso si voglia studiare la transizione senza risonanze.

Se si considerano solo i contributi a breve distanza, il calcolo del *branching ratio* porta a

$$BR(B_s^0 \rightarrow \mu^+\mu^-\gamma) \simeq 5 \cdot 10^{-9}. \quad (1.12)$$

Si può notare come questo numero sia dello stesso ordine di quello ottenuto per il decadimento senza il fotone. Infatti, sebbene questo processo avvenga ad un'ordine successivo, avendo un vertice in più, non subisce la soppressione per elicità grazie alla presenza del fotone nello stato finale.

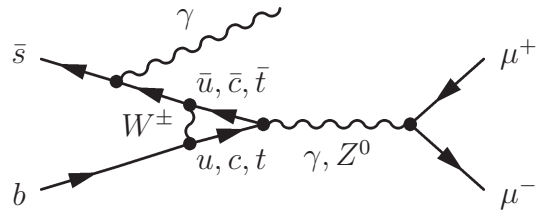
Alcuni calcoli con metodi *model independent*, portano a risultati simili sempre con $BR(B_s^0 \rightarrow \mu^+\mu^-\gamma) \sim 10^{-9}$ [11, 12].

Dal punto di vista sperimentale nelle misure di decadimenti puramente leptonici potrebbero contribuire i canali con fotoni di bassa energia non rivelabili dall'esperimento. Per questo motivo il decadimento $B_s^0 \rightarrow \mu^+\mu^-\gamma$, oltre che essere utile di per sé, sarà necessario per la valutazione del *branching ratio* del decadimento $B_s^0 \rightarrow \mu^+\mu^-$.

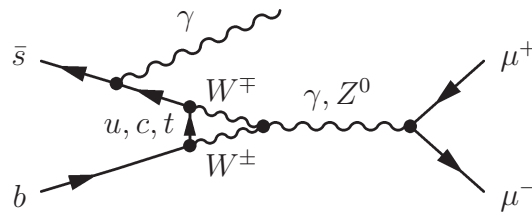
Se si considera lo spettro del fotone emesso, esso presenta una divergenza infrarossa rinormalizzabile; ciò è naturale in processi di questo tipo e si può mostrare che sommando i due contributi, $B_s^0 \rightarrow \mu^+ \mu^- \gamma$ e $B_s^0 \rightarrow \mu^+ \mu^-$, ogni divergenza si cancella e il risultato è di ordine finito [13].

Sulla ricerca di questo decadimento, recentemente è stato pubblicato un articolo dell'esperimento BABAR che ha posto il seguente limite al *branching ratio*: $BR(B_d^0 \rightarrow \mu^+ \mu^- \gamma) < 1.5 \cdot 10^{-7}$ [14].

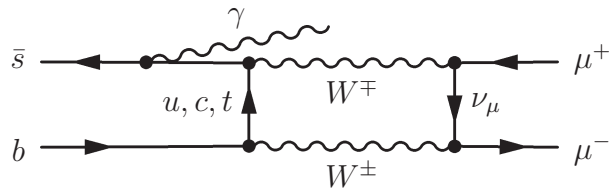
La maggiore statistica dell'esperimento LHCb dovrebbe portare ad una sensibilità migliore di questa; inoltre la maggiore energia nel centro di massa darà la possibilità di studiare i decadimenti del B_s^0 oltre che del B_d^0 ; in questo lavoro si è studiata la selezione di questo decadimento per la quale si rimanda al Capitolo 5.



(a)



(b)



(c)

Figura 1.3: Diagrammi di Feynman che contribuiscono al decadimento $B_s^0 \rightarrow \mu^+ \mu^- \gamma$ all'interno del Modello Standard

Capitolo 2

L'esperimento LHCb

2.1 LHC - Large Hadron Collider

Il Large Hadron Collider o LHC è l'acceleratore di particelle attualmente in costruzione al CERN. L'obiettivo è la produzione di collisioni protone-protone tramite le quali effettuare misure precise del Modello Standard ed evidenziare la presenza di eventuale nuova fisica.

Il LHC è situato nel tunnel circolare sotterraneo (27 km di circonferenza) che ospitava precedentemente l'acceleratore LEP ed è costituito da un acceleratore contenente due cavità per i due fasci di protoni che si muoveranno in senso inverso.

La macchina sarà operativa dalla primavera del 2008 e produrrà un'energia nel centro di massa delle collisioni pari a 14 TeV. La luminosità dei fasci sarà di $10^{34} \text{cm}^{-2} \text{s}^{-1}$.

L'acceleratore è costituito da due anelli a sincrotrone paralleli, che si incrociano in 4 punti di collisione dove avranno luogo i vari esperimenti: la struttura non è completamente circolare essendo costituita da 8 archi di 2.8 km intervallati da 8 sezioni in linea retta lunghe 540 metri. Riferendoci alla Figura 2.1, nei Punti 1, 2, 5 e 8 saranno installati gli esperimenti, i Punti 3 e 7 sono dedicati alla pulizia del fascio, il Punto 4 all'accelerazione e quello 6 all'estrazione.

La curvatura dei fasci è affidata a 1232 magneti dipoli superconduttori che produrranno un campo magnetico di 8.4 T lavorando a 1.9 K e 400 magneti

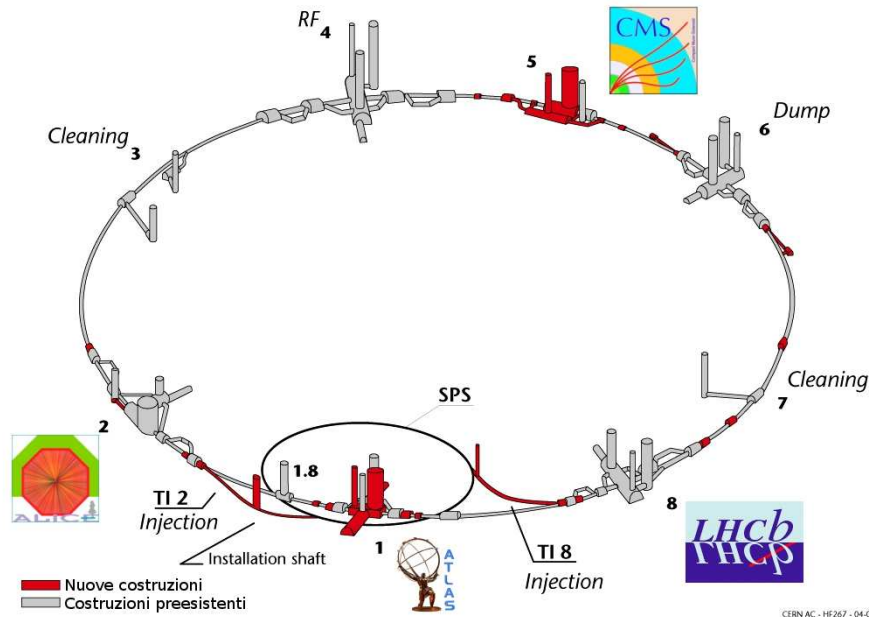


Figura 2.1: Schema delle strutture sotterranee del Large Hadron Collider

quadrupoli producenti un gradiente di 223 T m^{-1} . Sono inoltre presenti vari altri tipi di magneti per correggere errori di multipolo, focalizzare i fasci, guidarli nelle zone di interazione e aumentarne la luminosità per le collisioni.

Nel complesso LHC ospiterà cinque esperimenti: due esperimenti con multipli obiettivi fisici, ATLAS (nel Punto 1) e CMS (nel Punto 5), TOTEM (anch'esso nel Punto 5) per la misura della sezione d'urto totale dei protoni, ALICE (al Punto 2) che studierà interazioni tra ioni pesanti, e LHCb (al Punto 8) per lo studio della fisica dei B.

2.2 Collisioni protone-protone

Le interazioni tra due protoni dei fasci collidenti possono essere divise in due tipi:

- interazioni soffici, a piccolo angolo di scattering (dette anche *minimum bias*) che costituiscono la maggior parte degli eventi e che generalmente non sono interessanti;

Tabella 2.1: Sezioni d'urto per lo scattering p-p

Sezione d'urto totale	$\sigma_{tot} = 100mb$
Sezione d'urto inelastica	$\sigma_{in} = 55mb$
Produzione $c\bar{c}$	$\sigma_{c\bar{c}} = 3.5mb$
Produzione $b\bar{b}$	$\sigma_{b\bar{b}} = 500\mu b$

- interazioni dure, più rare, con grande trasferimento di impulso, che costituiscono gli eventi interessanti.

Alcune sezioni d'urto relative all'interazione p-p sono listate in Tabella 2.1.

All'interno di LHC i protoni saranno divisi in gruppi (*bunches*) distanti in tempo 25 ns. Gli scontri tra due gruppi di protoni in opposte direzioni (*bunch-crossing*) avverranno con una frequenza $f_{bc} = 10$ MHz, questo numero dipende dalla presenza di alcuni *bunches* privi di protoni.

Alla luminosità massima $\mathcal{L} = 10^{34} cm^{-2}s^{-1}$ si avranno circa 20 interazioni per *bunch-crossing*; l'esperimento LHCb invece lavorerà ad una luminosità $\mathcal{L} = 2 \cdot 10^{32} cm^{-2}s^{-1}$ portando a 0.3 il numero medio di interazioni per *bunch-crossing*.

2.2.1 Produzione di b

La produzione di coppie $b\bar{b}$ avviene secondo il diagramma generico in Figura 2.2(a); i tre contributi dominanti sono i diagrammi cosiddetti di *gluon splitting* (2.2(b)), *gluon fusion* (2.2(c)) e *flavour excitation* (2.2(d)).

La sezione d'urto alle energie di LHC è fortemente influenzata da incertezze teoriche [15] e varia in un intervallo tra 175 e 950 μb a seconda del valore di alcuni parametri. Una lieve influenza sulla sezione d'urto viene data dall'incertezza sulla massa del quark b nel polo, ma l'influenza maggiore la forniscono le incertezze sulle scale di rinormalizzazione e di frammentazione. Si assume dunque come valore della sezione d'urto il valore medio $\sigma_{b\bar{b}} = 500\mu b$. Questa sarà nota con maggior precisione una volta che LHC sarà in funzione.

Tabella 2.2: Frazioni di adroni b ottenute in decadimenti $Z^0 \rightarrow b\bar{b}$ e collisioni $p\bar{p}$ a $\sqrt{s} = 1.8$ TeV.

Adrone b	Frazione [%]
B^0, B^+	39.8 ± 1.0
B_s^0	10.4 ± 1.4
barioni b	9.9 ± 1.7

Data questa sezione d'urto ci si aspettano, alla luminosità media a cui prenderà dati LHCb, una media di 10^{12} coppie $b\bar{b}$ per anno (10^7 s). Ognuno dei due quark, alle energie di LHC, adronizzerà in mesoni B (B^0, B_s^0, B^+) o barioni (Es. Λ_b); le frazioni di adronizzazione sono riportate in Tabella 2.2 [16].

Come si è visto, il meccanismo fondamentale di produzione di coppie $b\bar{b}$ è dato dalla fusione di due gluoni. Le funzioni di distribuzione partoniche alle energie di LHC sono tali che ad *incontrarsi* durante le collisioni siano partoni (e quindi gluoni irradiati) con impulsi molto differenti, portando così ad una distribuzione pressoché piatta in rapidità.

Poiché la massa dei due quark b è molto inferiore all'energia del centro di massa, essi avranno un *boost* tale che le direzioni di volo dei due quark, e quindi degli adroni da loro prodotti, saranno correlate tra loro e piccate a bassi angoli polari (0 e π).

La geometria dell'perimento LHCb, progettato per lo studio della fisica dei b , deriva appunto da questo fatto.

2.3 Il rivelatore LHCb

Il rivelatore LHCb [17, 18] è uno spettrometro a singolo braccio posto nella regione *in avanti* delle interazioni pp . È dedicato allo studio della violazione della simmetria CP e di decadimenti e fenomeni rari nella fisica delle particelle dotate di b .

Per assolvere a questi scopi il rivelatore LHCb possiederà le seguenti caratteristiche:

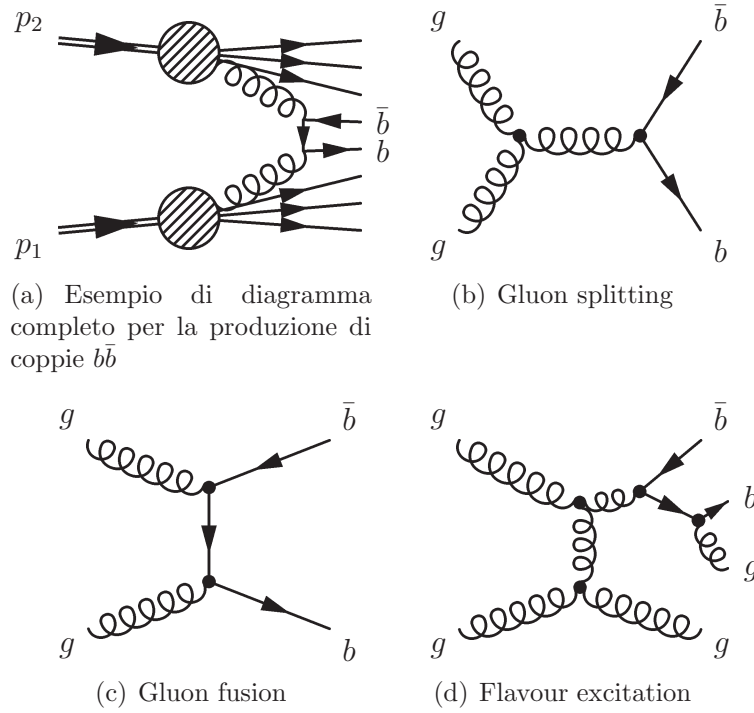


Figura 2.2: Diagrammi di Feynman per la produzione di coppie $b\bar{b}$. Esempio di diagramma completo (a) e contributi dominanti alla produzione della coppia (b), (c) e (d).

- alta efficienza di ricostruzione delle tracce;
- capacità di separare pioni da kaoni, e in genere un'ottima identificazione delle particelle;
- un'ottima risoluzione sul tempo di volo dei B ;
- alte efficienze di trigger sia per gli stati finali leptonici che adronici.

Il rivelatore è situato nel Punto 8 di LHC (Cfr. Figura 2.1), punto precedentemente occupato dall'esperimento DELPHI; LHCb è stato costruito in maniera tale da poter essere installato senza ulteriori lavori di ingegneria civile, le dimensioni sono quindi state limitate da quelle della caverna preesistente.

La geometria dell'esperimento è stata adattata allo studio dei B , prodotti fortemente *in avanti* in LHC come già detto; la sezione verticale del rivelatore è mostrata in Figura 2.3. Il sistema di coordinate di LHCb è destrorso con l'asse z che scorre dal punto di interazione verso le camere a muoni lungo la linea di fascio e con l'asse y che punta verso l'alto. Il magnete è orientato in maniera tale che le tracce cariche siano incurvate lungo il piano $x-z$ (detto piano orizzontale o piano curvante) e non siano incurvate nel piano $y-z$ (piano verticale o non-curvante).

La copertura angolare è tra 10 e 300 mrad nel piano $x-z$ e tra 10 e 250 mrad nel piano $y-z$. Questa equivale ad una copertura in pseudo-rapidità pari a $1.8 < \eta < 4.9$.

La luminosità di lavoro di LHCb ($\mathcal{L} = 2 \cdot 10^{32} \text{cm}^{-2} \text{s}^{-1}$), diversa da quella nominale di LHC ($\mathcal{L} = 10^{34} \text{cm}^{-2} \text{s}^{-1}$), è stata scelta in maniera tale da evitare il più possibile eventi con doppio vertice di interazione. Questo perché una buona misura di tempo proprio delle particelle e quindi una buona misura dei vertici secondari sarebbe inficiata dalla presenza di ulteriori vertici primari. Inoltre il sistema di *trigger* di LHCb, nella sua parte software di alto livello (HLT), è progettato per attivarsi alla presenza di vertici secondari: ulteriori vertici primari porterebbero questo ad attivarsi per eventi probabilmente non interessanti.

Il rivelatore è costituito dalle seguenti parti:

- il localizzatore di vertice (*vertex locator* o VELO);
- il Ring-Imaging Cherenkov Detector (RICH1) a monte;
- il Trigger Tracker (TT);
- il magnete;
- il sistema di tracciamento (T1,T2,T3);
- il Ring-Imaging Cherenkov Detector (RICH2) a valle;
- il sistema di *preshower* (SPD/PS);
- il calorimetro elettromagnetico (ECAL);
- il calorimetro adronico (HCAL);
- il sistema di rivelazione di muoni (M1, ..., M5).

che verranno discusse singolarmente di seguito.

2.3.1 Il VELO

Il Vertex Locator (VELO) [19] è posizionato nella zona di interazione dei protoni e fornirà una precisa ricostruzione del vertice primario di interazione e dei vertici secondari di decadimento degli adroni con *beauty* e *charm*.

È costituito da 25 stazioni di dischi traccianti di silicio con una segmentazione basata sulle coordinate r - ϕ (rispettivamente raggio e angolo azimutale in un sistema di coordinate polari cilindriche).

La risoluzione del rivelatore è pari a $42 \mu\text{m}$ nella direzione z e di $10 \mu\text{m}$ nella direzione perpendicolare al fascio. Si otterrà inoltre una risoluzione di $20 \mu\text{m}$ per la ricostruzione del parametro di impatto rispetto al vertice primario delle tracce a più alto impulso trasverso. La risoluzione per i vertici secondari varia tra 220 e $370 \mu\text{m}$ a seconda del canale di decadimento. Questo porterà, per esempio, ad una risoluzione di 40 fs sul tempo di vita per il decadimento $B_s^0 \rightarrow D_s^- \pi^+$ utile per la misura di Δm_s .

Le informazioni provenienti dal VELO verranno inoltre utilizzate per il trigger di alto livello (HLT).

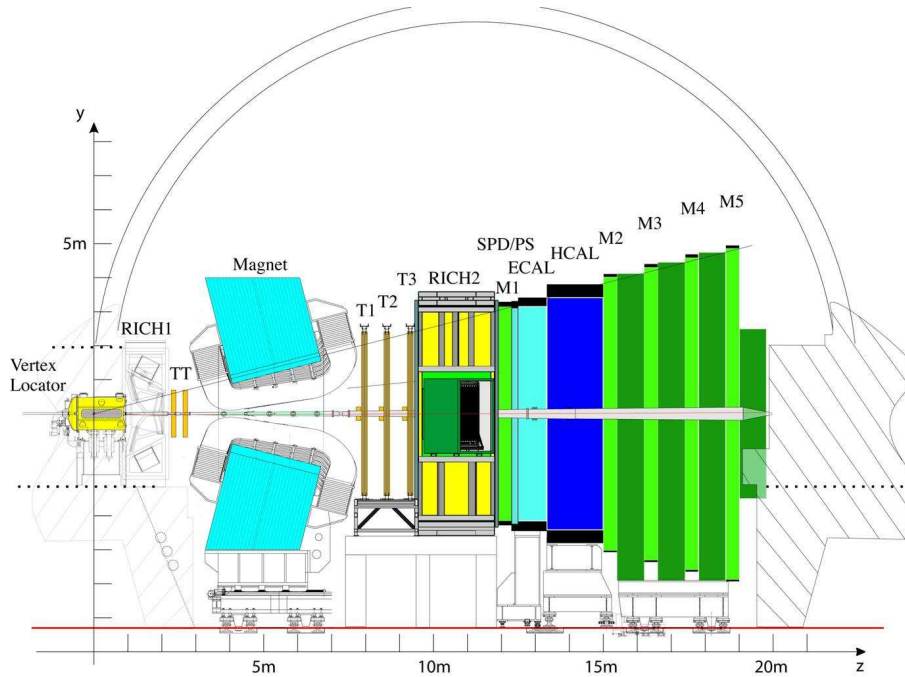


Figura 2.3: Il rivelatore LHCb: sezione verticale.

2.3.2 Il RICH

In LHCb sono presenti due RICH - Ring-Imaging Cherenkov Detectors [20], ossia rivelatori per la luce Cherenkov emessa dalle particelle cariche che li attraversano. Il maggior obiettivo di questi rivelatori è l'identificazione delle particelle e in particolare la separazione tra pioni e kaoni.

Il RICH rivela le immagini di anelli (sezioni di cono) di fotoni emessi da particelle cariche che attraversano uno strato di radiatore posto al suo ingresso. I fotoni vengono deviati da due serie di specchi (prima sferici poi piani) verso dei fotodiodi ibridi segmentati (HPD). Questi sono sensibili ai campi magnetici, ragion per cui sono schermati.

Per identificare particelle in un ampio intervallo di impulsi sono necessari tre diversi radiator: aerogel (indice di rifrazione $n = 1.03$) per i bassi impulsi, C_4F_{10} gassoso ($n = 1.0014$) per quelli intermedi e CF_4 ($n = 1.0005$) per i grandi impulsi.

Poiché vi è una forte correlazione tra l'angolo polare delle particelle e il loro impulso il RICH è diviso in due parti che coprono due diverse regioni di

angolo polare (e conseguentemente di impulso) delle particelle:

il RICH1, collocato a monte del magnete, subito dopo il VELO, usa sia aerogel che C_4F_{10} come radiatori. Essendo posto prima del magnete è capace di rivelare anche particelle che verranno portate fuori accettazione dal campo magnetico. Copre la regione di bassi impulsi (1-70 GeV) e di grandi angoli polari (30-300 mrad).

il RICH2, a valle del magnete, usa CF_4 come radiatore e copre la regione di alti impulsi (12-150 GeV) e bassi angoli polari (15-120 mrad).

Entrambi i rivelatori sono posti fuori del magnete, in regioni di basso campo magnetico, sia per evitare che le tracce siano incurvate mentre attraversano i radiatori, sia perché i fotodiodi sono sensibili al campo.

2.3.3 Il Trigger Tracker

Il Trigger Tracker (TT) [21] è collocato tra il RICH1 e l'ingresso del magnete. Esso servirà sia a delineare le tracce di basso impulso che, deviate dal magnete, non arriveranno alle altre stazioni di tracciamento, sia per il trigger di alto livello per assegnare informazioni di impulso trasverso alle tracce con alto parametro di impatto.

L'area attiva del rivelatore è composta di quattro piani di rivelatori a microstrip di silicio lunghe 33 cm e con un passo di 198 μm . I quattro piani sono divisi in due coppie separate da una distanza di 30 cm. Il primo e il quarto piano hanno strip di lettura verticali, mentre il secondo e il terzo hanno strip con coordinate stereo cioè ruotate di più e meno 5° rispetto alla verticale.

2.3.4 Il magnete

Il magnete dipolare [22] è collocato vicino al punto di interazione, per poter essere di dimensioni contenute, ma dopo uno schermo di ferro che protegge il VELO e il RICH1 dal campo magnetico.

L'apertura del magnete è di 300 mrad nel piano orizzontale e di 250 in quello verticale.

È composto da 50 tonnellate di cavi di alluminio conduttore (9 km in totale) in due bobine trapezoidali piegate a 45 gradi rispetto al piano orizzontale, incluse in un giogo che incanala il flusso di ritorno del campo, costituito da 120 kt di piatti in acciaio.

Il campo è orientato verticalmente, cosicché le tracce vengono incurvate lungo il piano x - z . Ha un'intensità massima di 1.1T ed un campo integrato $\int \vec{B} \cdot d\vec{l} = 4 \text{ T}\cdot\text{m}$ in media per tracce originatesi vicino al vertice primario, questo per consentire una risoluzione del 0.4% per la misura di impulsi fino a 200 GeV.

2.3.5 Il Tubo del fascio

Dopo l'uscita dal VELO il fascio dell'LHC è protetto da un tubo per il fascio (*beam pipe*) appositamente costruito.

È diviso in varie parti: la prima è costituita da uno cono di berillio di 25 mrad di apertura e 1 mm di spessore; la seconda e terza parte sono costituite di una lega 40%-60% Al-Be di apertura di 10 mrad. L'ultima parte è costituita da un cono di apertura 10 mrad di acciaio inossidabile.

2.3.6 Il sistema di tracciamento

Il sistema di tracciamento [21, 23] fornisce misure di posizione delle tracce cariche e consente, insieme ai punti forniti dal VELO e dal Trigger Tracker di misurare l'impulso delle particelle.

Il sistema è costituito da tre stazioni traccianti (T1, T2, T3) ognuna delle quali misura la coordinata x della particella e due coordinate *stereo* a più e meno 5° dalla verticale.

Il Tracker è distinto in due parti, interna (*Inner Tracker*) ed esterna (*Outer Tracker*), questo perché la densità di tracce varia circa come $1/r^2$ dove r è la distanza dall'asse z e sono quindi necessarie diverse risoluzioni.

L'*Inner Tracker* copre la parte più interna delle stazioni T1, T2, T3, in una regione $40 \times 120 \text{ cm}^2$ (altezza \times larghezza) attorno al tubo del fascio. Esso è costituito da traccianti in silicio, simili a quelli del Trigger

Tracker: con fogli di microstrip di silicio di spessore $340 \mu\text{m}$ e passo tra le strip di $230 \mu\text{m}$. Possiede una risoluzione spaziale di $70 \mu\text{m}$.

L'*Outer Tracker* ricopre il resto delle stazioni di tracciamento ed è costituito da camere a deriva dette *straw-tubes*, esse hanno un diametro di 5 mm e pareti di $75 \mu\text{m}$: per ottenere la risoluzione in impulso pari a $\delta p/p = 0.4\%$ esse devono avere una buona misura della posizione nel piano (curvante) $x-z$, motivo per cui hanno la struttura con i tre piani a 0° e $\pm 5^\circ$ sopra descritta.

2.3.7 I Calorimetri

Il sistema di calorimetri [24] permette di identificare elettroni, fotoni e adroni e di misurarne posizione e impulso: queste informazioni sono utilizzate anche nel trigger di livello 0 dell'apparato. Il design del sistema è influenzato fortemente da quest'ultima esigenza.

L'intero sistema è costituito da un rivelatore a scintillazione a pad (*scintillator pad detector* o SPD), un *preshower*, un calorimetro elettromagnetico e uno adronico. L'SPD serve ad identificare le tracce cariche che entrano nel calorimetro. Il *preshower*, costituito da 12 mm di piombo seguiti da un secondo SPD, serve a discriminare i fotoni e i pioni dagli elettroni in maniera rapida, per avere un'informazione utilizzabile già al livello 0 di trigger.

Il calorimetro elettromagnetico ECAL

Il calorimetro elettromagnetico (ECAL) rivela elettroni e fotoni per mezzo dell'identificazione delle cascate elettromagnetiche provocate da queste particelle in interazione con il materiale. La sua lunghezza totale di radiazione è di $25 X_0$. È costruito secondo la tecnologia *Shashlik* ed è costituito di piani di 2 mm di piombo intervallati da piani di 4 mm di scintillatori. La risoluzione ottenibile è:

$$\frac{\sigma(E)}{E} = \frac{10\%}{\sqrt{E}} \oplus 1.5\%$$

Il calorimetro adronico HCAL

HCAL è un calorimetro adronico a campionamento che utilizza *mattonelle* di ferro e di scintillatore, rispettivamente come assorbitore e materiale attivo. Si hanno 4 mm di scintillatore intervallati da 16 mm di ferro, per una lunghezza totale lungo z pari a 1.6 m o $5.6 \lambda_I$. La risoluzione ottenuta è pari a

$$\frac{\sigma(E)}{E} = 80\% \sqrt{E} \oplus 10\%$$

2.3.8 Il sistema per muoni

Il rivelatore per muoni [25, 26, 27] è posto all'estremo del rivelatore LHCb: esso ha la funzione di rivelare i muoni, uniche particelle cariche a poter attraversare i calorimetri quasi senza interagire. Muoni di alto impulso trasverso vengono prodotti nei decadimenti dei B, per questo motivo questo rivelatore è fondamentale per il livello 0 del trigger. Oltre a ciò il sistema per muoni sarà necessario per l'identificazione *offline* dei muoni, utile per la selezione precisa di vari decadimenti.

Il sistema è costituito da cinque stazioni di rivelatori (M1, ..., M5), la prima delle quali, M1, è posta davanti al calorimetro elettromagnetico mentre le altre sono poste dopo il calorimetro adronico.

Ogni stazione è divisa in regioni con *pad* di dimensioni differenti che aumentano con la distanza dall'asse del fascio, le regioni sono dette R1, ..., R4. I rivelatori sono tutti costituiti da camere proporzionali multifilo (MWPC) ad eccezione della regione R1 del rivelatore M1 che invece è stata equipaggiata con rivelatori a tripla GEM, per il loro migliore comportamento nelle zone ad alto flusso. Le stazioni M2, ..., M5 sono intervallate da filtri di ferro che servono ad attenuare le particelle, non muoni, che hanno oltrepassato i calorimetri.

2.3.9 Il trigger

La frequenza di eventi in LHCb si è detto, è dell'ordine di 10 MHz, questa ovviamente dev'essere ridotta a qualche centinaia di Hz, frequenza alla quale

gli eventi possono essere scritti su memoria per poter essere analizzati successivamente. La selezione di questi eventi interessanti è effettuata dal sistema di trigger di LHCb [28].

Esso è diviso in un trigger hardware, di livello zero (L0), e uno software, di alto livello (HLT).

Il trigger L0 ha lo scopo di ridurre il rate di eventi da 10 MHz a 200 kHz, rate accessibile al trigger software che richiede tempi maggiori. Esso ricostruisce le particelle di p_T massimo, utilizzando le informazioni del calorimetro e del *muon system* e valuta la possibilità che provengano da un B . Inoltre richiede che l'evento sia di bassa molteplicità.

I dati analogici del rivelatore vengono quindi digitizzati, e immagazzinati in un *buffer* letto dal HLT. Esso valuta l'informazione di tutto il rivelatore e seleziona gli eventi di interesse in base ad un algoritmo scritto in C++. Il suo output sarà ad un rate di 1 kHz, al quale gli eventi verranno scritti su memoria.

2.3.10 Il sistema di *computing* di LHCb

Come si è visto l'esperimento LHCb produrrà grandi quantitativi di dati, che dovranno essere digitizzati, ricostruiti e analizzati. Inoltre si stanno ancora studiando le performance e le caratteristiche del rivelatore per ottenere misure più precise.

Il software, comune ai vari livelli di lavoro, dev'essere contemporaneamente stabile e flessibile per adattarsi alle caratteristiche dell'esperimento per un periodo di 10-20 anni. Queste e altre esigenze hanno imposto la struttura del software di LHCb [29].

Al centro del sistema di computazione vi è un'architettura in grado di interfacciarsi con i vari livelli di computazione dell'esperimento e di tenere sotto controllo i cambiamenti tecnici mantenendo ferma la struttura globale del software.

L'architettura che si occupa di questo è chiamata GAUDI, essa è basata su una struttura Object Oriented scritta in linguaggio C++. Scopo principale di GAUDI è definire un set di strutture e interfacce che siano comuni alle

varie applicazioni, mantenendo queste disaccoppiate tra loro. Inoltre questi servizi sono stati pensati in maniera indipendente dalla loro implementazione effettiva, per permettere semplicità e robustezza al software, nascondendo i dettagli tecnici all'utilizzatore finale.

GAUDI dunque fa da struttura portante per i vari programmi, per le varie fasi del processamento dati, elencati di seguito:

GAUSS: è programma di simulazione di eventi basato su GEANT4; esso può generare degli eventi simulati (Monte Carlo) e la loro interazione fisica con il rivelatore.

BOOLE: è la fase finale della simulazione di eventi, riproduce la risposta del rivelatore alle interazioni generate con GAUSS; il suo output ha la stessa forma dei dati reali provenienti dal rivelatore.

BRUNEL: ricostruisce gli eventi, a partire dalla risposta del rivelatore, producendo entità fisiche (come tracce) per produrre set di dati da analizzare;

DAVINCI: è il programma di analisi dati di LHCb; comprende la ricostruzione di vertici, l'identificazione di particelle e algoritmi di selezione.

Oltre ai programmi suddetti vi sono vari altri programmi sia di analisi interattiva che di visualizzazione.

EvtGen

Per la simulazione della fisica dei decadimenti dei B viene usato da GAUDI il pacchetto EvtGen [30]. In particolare EvtGen permette di generare decadimenti a cascata e decadimenti con violazione di CP, utili appunto per lo studio della fisica dei B .

Esso sfrutta uno sviluppo delle ampiezze di decadimento in ampiezze di elicità, che accoppiano stati iniziali e finali di spin definito. Questo permette di riprodurre efficacemente le distribuzioni angolari e le dipendenze temporali nei decadimenti in sequenza dei B .

È possibile inoltre la generazione *Phase Space* dei decadimenti, che considera solo lo spazio delle fasi disponibile per lo stato finale.

Capitolo 3

Metodi multidimensionali di selezione degli eventi

La selezione degli eventi, solitamente e storicamente, viene attuata scegliendo per ogni variabile a disposizione (cinematica o dinamica) una regione di confidenza alla quale si crede appartenga il segnale. Questo metodo che consiste appunto nell'applicare dei *tagli*¹ sulle variabili, ha vari pregi: oltre alla semplicità, ha un chiaro significato fisico ed è di più facile gestione per il calcolo delle sistematiche.

Tuttavia questo metodo ha diversi difetti. A livello puramente teorico è chiaro come non sfrutti tutta l'informazione contenuta nei dati raccolti. Infatti un evento fisico è caratterizzato da una serie di innumerevoli informazioni correlate tra loro. Quando si analizzano le distribuzioni di alcune variabili, invece, si proietta lo spazio delle informazioni in una particolare direzione; questa proiezione perde traccia delle correlazioni degradando così parte dell'informazione.

I metodi multidimensionali (o multivariati), invece, analizzano i dati fisici tentando di recuperare anche la parte dovuta al carattere a più dimensioni dell'informazione. Questo viene effettuato o decorrelando antecedentemente le variabili, o scegliendo assi di proiezione più vantaggiosi, o ancora tramite metodi di scelta non lineari.

In questo lavoro si sono considerati vari metodi, che verranno esposti qui

¹Di seguito questo metodo sarà detto *tagli rettangolari* perché stabilisce dei limiti netti alle variabili.

di seguito, in particolare per l'applicazione alla selezione di eventi rari, per i quali è ancora più cruciale lo sfruttamento al massimo grado dell'informazione contenuta nei dati.

3.1 Metodi multidimensionali

3.1.1 Likelihood

Dato un evento caratterizzato da un numero n_{var} di variabili x_i , e data una (k -esima) categoria tra varie tra le quali classificare l'evento, si definisce Likelihood \mathcal{L}^k (in italiano *verosimiglianza*) la seguente funzione:

$$\mathcal{L}^k = \prod_i^{n_{var}} p_i^k(x_i) \quad (3.1)$$

cioè la produttoria delle probabilità che l' i -esima variabile dell'evento possa appartenere ad un evento della categoria k . Le p_i^k sono in generale delle funzioni di distribuzione di probabilità (PDFs - *Probability Distribution Functions*) e, affinché la likelihood stessa possa essere interpretata come una probabilità, devono essere normalizzate a uno, cioè:

$$\int_D p_i^k(x_i) dx_i = 1 \quad \forall k, i \quad (3.2)$$

dove D è l'intero dominio della variabile x_i .

La definizione viene spesso usata nella seguente forma logaritmica (*log-Likelihood* o $\ell\ell$) per utilizzare somme al posto di prodotti:

$$\ln \mathcal{L}^k = \sum_i^{n_{var}} \ln p_i^k(x_i). \quad (3.3)$$

È chiaro come questa variabile abbia la capacità di discriminare tra diverse categorie di eventi. In particolare se, come spesso accade, si hanno due categorie tra le quali classificare un evento si utilizza solitamente il rapporto tra le Likelihood come variabile discriminante, questo ovviamente si trasforma nella differenza tra le *log-Likelihood*: ci riferiremo a ciò con il termine

Delta-log-Likelihood.² Spesso inoltre nella selezione di eventi, come anche nel nostro caso, si usa un rapporto delle likelihood normalizzato, cioè detta \mathcal{L}_S (\mathcal{L}_B) la likelihood per un evento di essere segnale (fondo), si definisce il rapporto delle likelihood normalizzato come:

$$\rho_{\mathcal{L}} = \frac{\mathcal{L}_S}{\mathcal{L}_S + \mathcal{L}_B} \quad (3.4)$$

Il vantaggio è che il suo valore è compreso tra 0 e 1, dove vicino a 1 saranno gli eventi appartenenti al segnale e a zero quelli del fondo.

Trasformazione della variabile di output

La definizione di rapporto di likelihood suddetta, fornisce un ottimo modo per discriminare eventi di segnale e fondo. Una distribuzione di rapporto di likelihood tipica è quella mostrata in Figura 3.1 come esempio, come si può vedere le due categorie sono ben separate, e pur tuttavia eventi di una categoria si estendono sin sotto il picco dell'altra. Nel caso di selezione di eventi rari spesso è conveniente utilizzare selezioni più *pure* sebbene meno efficienti (approfondiremo questo nel §3.2.3); per ottenere questo si è dimostrata utile una trasformazione del rapporto di likelihood. La trasformazione in questione è la seguente:

$$\rho' = -\frac{1}{\tau} \ln \left(\frac{1}{\rho} - 1 \right) \quad (3.5)$$

dove ρ è il rapporto definito in (3.4) e τ un numero che definisce solo la scala e non influenza la trasformazione. La funzione così definita non è altro che la funzione sigmoidea inversa. Si noti che ρ' non sarà più compreso tra zero e 1.

Il risultato di questa trasformazione è mostrato in Figura 3.2: come si può vedere le due popolazioni risultano meno distinte, ma le *code* delle distribuzioni sono nitide, e possono essere usate per selezioni più *pulite*.

²Si noti che spesso ci si riferisce a questa variabile parlando di DLL, come ad esempio nel caso dell'identificazione delle particelle, che viene effettuata proprio costruendo delle *Delta-Log-Likelihood* a partire dalle variabili che il rivelatore mette a disposizione.

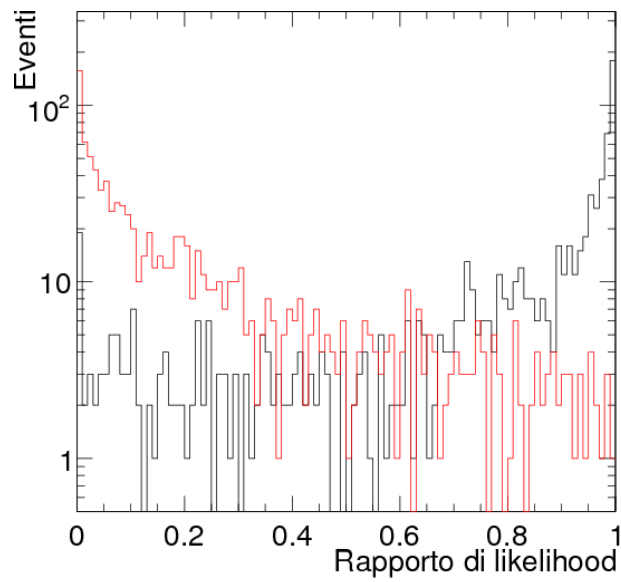


Figura 3.1: Esempio di distribuzione di Rapporto di Likelihood per segnale (in nero) e fondo (in rosso).

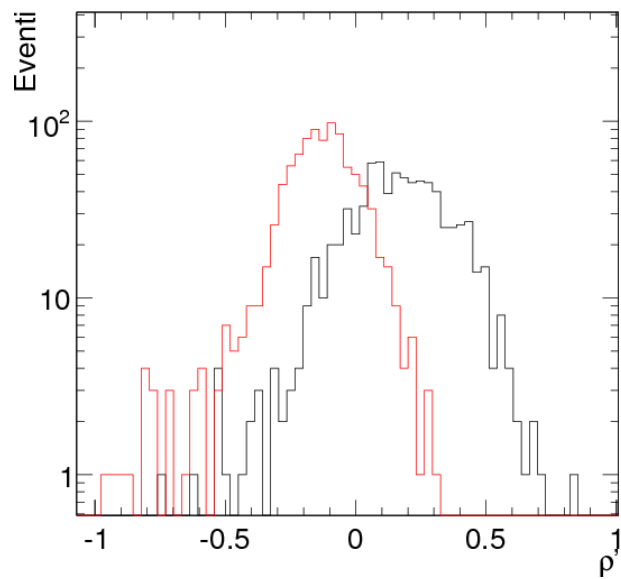


Figura 3.2: Esempio di distribuzione di rapporto di Likelihood, trasformato secondo la (3.5) per segnale (in nero) e fondo (in rosso).

3.1.2 Decorrelazione delle variabili

Un problema della definizione di *Likelihood*, ma anche di altri metodi basati sulle probabilità è che la probabilità congiunta delle n_{var} variabili, è il prodotto delle singole PDF solamente se le variabili sono indipendenti, cioè:

$$P(A \cap B) = P(B|A)P(A)$$

$$P(A \cap B) = P(A)P(B) \iff A \text{ e } B \text{ sono indipendenti}$$

La definizione di *Likelihood* rende dunque conto della probabilità che un evento appartenga al segnale piuttosto che al fondo solo in maniera parziale, se non scorretta, se le variabili usate sono correlate.

Per ovviare a questo problema si utilizzano metodi di *decorrelazione* delle variabili.

Quando le variabili sono correlate solo linearmente, il metodo per decorrelarle è relativamente semplice e verrà discusso qui di seguito; per correlazioni non lineari esistono metodi di decorrelazione, ma sono spesso di difficile implementazione e controllo.

Decorrelazione lineare

Dato un campione di N eventi, sia \mathbf{x}_n il vettore delle variabili che caratterizzano l' n -esimo evento e sia Σ la matrice di covarianza, i cui elementi sono costruiti a partire dagli eventi del campione come

$$\Sigma_{ij} = \frac{1}{N-1} \sum_{n=1}^N (x_n^i - \bar{x}^i)(x_n^j - \bar{x}^j) \quad (3.6)$$

dove \bar{x}^i è la media nel campione dell' i -esima variabile. Si consideri la trasformazione di variabili:

$$\mathbf{y} = \mathbf{C}^{-1}\mathbf{x} \quad (3.7)$$

dove \mathbf{C} è una matrice simmetrica tale che

$$\mathbf{C}^\dagger \mathbf{C} = \mathbf{C} \mathbf{C} = \Sigma \quad (3.8)$$

tramite la quale si ricava la seguente:

$$\mathbf{x}^\dagger \Sigma^{-1} \mathbf{x} = \mathbf{y}^\dagger \mathbf{C}^\dagger \Sigma^{-1} \mathbf{C} \mathbf{y} = \mathbf{y}^\dagger \mathbf{y} \quad . \quad (3.9)$$

Si ottiene così un nuovo vettore di variabili \mathbf{y} le cui componenti sono decorrelate tra loro. La decorrelazione è in realtà esatta e facilmente dimostrabile nel caso di variabili distribuite gaussianamente e linearmente correlate. In questo caso infatti la distribuzione iniziale delle variabili sarebbe

$$f(\mathbf{x}) = f(x_1, x_2, \dots, x_{n_{var}}) = \frac{1}{(2\pi)^{n_{var}/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\dagger \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})} \quad (3.10)$$

dove $\boldsymbol{\mu}$ è semplicemente il vettore delle medie delle variabili. La distribuzione delle variabili trasformate invece è facile dimostrare che sarebbe

$$g(\mathbf{y}) = g(y_1, y_2, \dots, y_{n_{var}}) = \frac{1}{(2\pi)^{n_{var}/2}} e^{-\frac{1}{2}(\mathbf{y}-\bar{\mathbf{y}})^\dagger (\mathbf{y}-\bar{\mathbf{y}})} = \quad (3.11)$$

$$= \prod_{i=1}^{n_{var}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y_i - \bar{y}_i)^2} \quad (3.12)$$

cioè una distribuzione gaussiana multidimensionale fattorizzabile nelle singole distribuzioni, gaussiane, delle componenti.

3.1.3 Approssimazione di proiezione e correlazione (PCA)

Il presente metodo, descritto in [31], utilizza delle approssimazioni di proiezione per calcolare le distribuzioni di probabilità.

Si consideri un set di N variabili distribuite con una densità di probabilità di forma arbitraria $\mathcal{P}(\mathbf{x})$. Ora si consideri una distribuzione gaussiana N -dimensionale:

$$f(\mathbf{x}) = f(x_1, x_2, \dots, x_N) = \frac{1}{(2\pi)^{N/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\dagger \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})} \quad (3.13)$$

essa sarà, ovviamente, una rozza approssimazione di $\mathcal{P}(\mathbf{x})$.

Tuttavia si può sempre effettuare una trasformazione di variabili dalle x_i con distribuzioni di forma generica $p_i(x_i)$ a $y_i(x_i)$ tali che le distribuzioni delle y_i siano gaussiane, secondo la seguente formula:

$$y_i(x_i) = \sqrt{2} \operatorname{erf}^{-1}(2F_i(x_i) - 1) \quad (3.14)$$

dove erf^{-1} è la funzione d'errore inversa e $F_i(x_i)$ è la funzione cumulativa normalizzata della variabile x_i , cioè:

$$F_i(x_i) = \frac{\int_{x_{i,min}}^{x_i} p(x') dx'}{\int_{x_{i,min}}^{x_{i,max}} p(x') dx'} \quad (3.15)$$

Ora, in generale la distribuzione N -dimensionale delle y_i non sarà una gaussiana, sebbene lo siano le sue proiezioni. Tuttavia, assumendo che lo sia in prima approssimazione, si può costruire la probabilità come:

$$P(\mathbf{x}) = \frac{1}{|\Sigma|^{1/2}} e^{-\frac{1}{2}\mathbf{y}^\dagger(\Sigma^{-1}-\mathbf{I})\mathbf{y}} \prod_{i=1}^N p(x_i) \quad (3.16)$$

Infatti

$$\begin{aligned} P(\mathbf{x}) &= \frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}\mathbf{y}^\dagger(\Sigma^{-1})\mathbf{y}} = \\ &= \frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}\mathbf{y}^\dagger(\Sigma^{-1}-\mathbf{I})\mathbf{y}} \prod_{i=1}^N e^{-\frac{1}{2}y_i^2} \end{aligned} \quad (3.17)$$

Ora considerando l'ultima parte della (3.17) per una generica y , e sostituendovi la (3.14) si ha:

$$\begin{aligned} e^{-\frac{1}{2}y^2} &= \exp \left\{ -\frac{1}{2} \left(\sqrt{2} \operatorname{erf}^{-1}(2F(x) - 1) \right)^2 \right\} \\ &= \frac{\sqrt{\pi}}{2} \frac{d}{dx} \operatorname{erf} \left(\operatorname{erf}^{-1}(2F(x) - 1) \right) \\ &= \frac{\sqrt{\pi}}{2} \frac{d}{dx} \left(2 \frac{\int_{x_{min}}^x p(x') dx'}{\int_{x_{min}}^{x_{max}} p(x') dx'} - 1 \right) \\ &= \frac{\sqrt{\pi}}{\left(\int_{x_{min}}^{x_{max}} p(x') dx' \right)} p(x) \end{aligned} \quad (3.18)$$

Sostituendo la (3.18) per ogni x_i nella (3.17) si ha la (3.16).

A questo punto quindi si può costruire il rapporto di probabilità come:

$$\mathcal{R} = \frac{P_S(\mathbf{x})}{P_S(\mathbf{x}) + P_B(\mathbf{x})} \quad (3.19)$$

e utilizzarlo con il significato descritto per l'analogo della *likelihood*.

Si noti che l'approssimazione PCA è esatta nei seguenti casi:

- per distribuzioni con variabili non correlate (per cui $\Sigma \simeq \mathbf{I}$);
- per distribuzioni gaussiane modificate da trasformazioni unidimensionali di variabili.

Per molti altri casi l'approssimazione è comunque buona. Tuttavia poiché non è escluso che l'approssimazione non sia ottimale, è opportuno controllarla con opportuni metodi (Es.: test del χ^2).

3.1.4 Fisher

Il metodo del discriminante lineare di Fisher [32] costituisce un modo per selezionare gli eventi tramite una combinazione opportuna delle variabili.

La base di questo metodo è la ricerca di un asse nello spazio delle variabili, lungo la quale le proiezioni degli eventi di diverse categorie abbiano medie il più possibili separate e distinte (cioè con piccole deviazioni standard).

Si può illustrare il metodo in maniera semplice come segue. Date, per semplicità, due categorie di eventi, indicate da $k = 1, 2$, e date n variabili x_i^k appartenenti a questi eventi, si costruisce una funzione lineare delle variabili come segue:

$$F^k = \sum_{i=1}^n \lambda_i x_i^k \quad (3.20)$$

Per ogni funzione F così costruita si può costruire la differenza tra le medie delle due categorie:

$$D = \sum_{i=1}^n \lambda_i d_i = \boldsymbol{\lambda} \mathbf{d} \quad (3.21)$$

mentre la deviazione standard all'interno di una categoria sarà proporzionale a:

$$S = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j S_{ij} = \boldsymbol{\lambda} \mathbf{S} \boldsymbol{\lambda} \quad (3.22)$$

dove S_{ij} è la covarianza della i - e j -esima variabile, e quindi \mathbf{S} è la matrice di covarianza. Allora i coefficienti λ_i che massimizzano la distanza pesata in deviazioni standard tra le due popolazioni saranno quelli che massimizzano il rapporto D^2/S . Tramite una semplice ottimizzazione si può mostrare che la soluzione sarà costituita dal vettore di coefficienti $\boldsymbol{\lambda}$ soddisfi l'equazione

$$\mathbf{S} \boldsymbol{\lambda} = \mathbf{d} \quad (3.23)$$

Costruiti in questo modo i *coefficienti di Fisher* λ_i^* , si può costruire per un generico evento t il *discriminante di Fisher* come segue:

$$y_{F_i}(t) = \sum_{i=0}^n \lambda_i^* x_i(t) \quad (3.24)$$

Vi è da notare che sebbene questo metodo elimini di fatto le correlazioni lineari non può rimuovere correlazioni d'ordine superiore.

3.1.5 Reti neurali artificiali

Per reti neurali artificiali (ANN) si intende in generale delle simulazioni di una rete di *neuroni* connessi tra loro; ogni neurone produrrà una particolare risposta per ogni dato input.

La potenza delle reti neurali deriva dal poter assegnare ad ogni neurone una funzione di risposta non lineare, rendendo così non lineare l'intera rete; possono quindi essere usate per lo studio di problemi complessi che in ambito informatico sono altrimenti difficilmente modellizzabili.

Il funzionamento di una rete neurale è semplice: l'input è fornito ad una serie di neuroni in ingresso, questi, secondo la loro funzione di risposta, producono un output che andrà ad altri neuroni, in una rete che può essere resa complicata a piacere, e che è generalmente nascosta all'utilizzatore. Le comunicazioni dei neuroni continuano fino ad una serie di neuroni finali che forniscono l'output all'utente. Se la rete viene inizialmente *istruita* a fornire un output ben definito con una serie di problemi campione noti, in seguito può essere utilizzata per produrre decisioni su problemi simili di soluzione ignota.

Il metodo più semplice per produrre una rete neurale è il cosiddetto *multilayer*, o multistrato: dati n neuroni, in principio sarebbero possibili n^2 connessioni; queste possono essere ridotte dividendo i neuroni in strati e permettendo connessioni solo tra uno strato e l'altro.

Ogni neurone, detto anche *perceptrone*, è connesso agli altri attraverso una serie di pesi w , e fornisce un singolo *output* dati in input i valori di uscita di altri neuroni e i pesi. A questa combinazione applica inoltre una *funzione di attivazione* che può essere anche non lineare, matematicamente:

$$y_{j,k} = \varphi \left(\sum_{i=0}^n w_{j,k}^{j-1,i} y_{j-1,i} + b \right) \quad (3.25)$$

dove $y_{j,k}$ è l'output del k -esimo perceptrone della j -esimo *layer*, $w_{j,k}^{j-1,i}$ i pesi reciproci e b un *bias* che può essere sempre introdotto. Come funzione di risposta $\varphi(x)$ solitamente viene scelta o la funzione sigmoidea $\frac{1}{1+e^{-x}}$ o la $\tanh(x)$, questo per la proprietà di essere lineari all'origine e di saturare

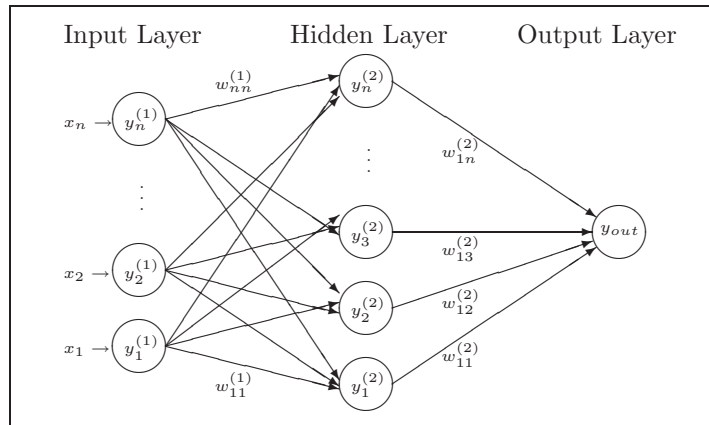


Figura 3.3: Schema di una rete neurale *multilayer* dotata di uno strato di input, uno nascosto e uno di output.

velocemente lontano da questa, in maniera tale da poter riprodurre problemi sia leggermente che fortemente non lineari.

Se consideriamo ora una rete neurale multilayer (come quella mostrata in Figura 3.3) dotata di uno strato di input, uno di output e uno nascosto, assumendo per semplicità che gli strati di input e output abbiano funzioni di attivazione lineari, l'output dell'intera rete sarà:

$$y_{out} = \sum_{j=1}^{n_h} y_j^{(2)} w_{j1}^{(2)} = \sum_{j=1}^{n_h} \varphi \left(\sum_{i=1}^{n_{in}} x_i w_{ij}^{(1)} \right) w_{j1}^{(2)} \quad (3.26)$$

dove n_{in} e n_h sono il numero di neuroni di input e nascosti, e w_{ij} i pesi tra i neuroni dell' i -esimo e j -esimo neurone di due strati. La scelta dei pesi viene fatta durante la fase di *training* descritta nel §3.2.

3.1.6 Decision Trees

I *Decision Trees* sono metodi di identificazione tra categorie che attuano una scelta in base ad un percorso di decisioni realizzate una variabile per volta; il percorso prosegue finché non viene soddisfatto un certo criterio. Lo spazio delle variabili viene così suddiviso in una serie di regioni classificate come appartenenti ad una certa categoria o ad un'altra.

Questo metodo di selezione, formalmente, è simile ai *tagli rettangolari*, tuttavia, mentre questi sono capaci di selezionare solamente un'iper-parallelepipedo nello spazio delle variabili, i *decision trees* sono capaci di selezionare svariati iper-parallelepipedi, dando luogo a performance migliori.

Un sistema comunemente usato per migliorare ulteriormente i risultati ottenuti con questo metodo è il cosiddetto *boosting*: questo consiste nel costruire vari *decision trees* dallo stesso campione iniziale ma con differenti pesi per gli eventi; questi *trees* sono quindi combinati per formare un metodo che decida la categoria di un evento in base a un *voto a maggioranza*.

Un difetto di questo metodo è il fatto che, spesso, risulti troppo *cucito addosso* ai dati che si usano per metterlo a punto; mentre non è detto che abbia le stesse performance su altri dati. Di questo fenomeno detto *overtraining* ci occuperemo nel §3.2.2.

3.2 Training, Ottimizzazione e criterio di Punzi

L'utilizzo di tutti i metodi sopra descritti richiede essenzialmente tre fasi:

- una fase di *training*;
- una fase di test;
- l'applicazione vera e propria.

3.2.1 Training

La fase di *training* consiste nell'istruire, nel mettere a punto i metodi scelti per il problema in questione. Il *training* è diverso da metodo a metodo, tuttavia il fine comune è impostare i parametri del sistema scelto al fine di riconoscere le categorie da distinguere: questo può essere effettuato, naturalmente, se si ha già una conoscenza (diretta o indiretta) delle due categorie e delle distribuzioni delle variabili da utilizzare.

In fisica delle alte energie questo problema viene affrontato spesso, come anche in questo caso, tramite simulazioni Monte Carlo dei dati che bisogna trattare.

In ogni caso alla fine del *training* si avrà una variabile di output distribuita in diversi modi per segnale e fondo e su questa si attuerà la scelta della zona da selezionare a seconda di vari criteri: questi possono essere criteri di efficienza, purezza o più raffinati come descriveremo in §3.2.3.

Dati dunque una serie di eventi noti si effettua il *training* nei seguenti modi:

- le funzioni di likelihood, definite in (3.1), e le funzioni del metodo PCA vengono costruite a partire dalle distribuzioni delle variabili di questi eventi. Da queste si costruirà il *likelihood ratio* ed eventualmente la sua versione trasformata secondo (3.5), e ciò costituirà l'output.
- Il *training* del metodo di Fisher consiste nella risoluzione, a partire dai dati, dell'equazione (3.23), questa fornirà i *coefficienti di Fisher* che meglio si adattano ai dati in questione.
- Per le reti neurali il concetto di *training* coincide con il significato della parola (allenamento), si tratta infatti di fornire come input della rete neurale gli eventi, e di richiedere (forzatamente) un certo output. Questo si ottiene aggiustando i pesi tra i neuroni; il metodo comunemente usato è detto *backpropagation*, propagazione a ritroso: si calcola l'output per ogni evento e lo si confronta (tramite una funzione d'errore) all'output desiderato, i pesi che minimizzano questa funzione d'errore saranno quelli desiderati.
- Il training dei *decision trees* costruisce il metodo per *nodi*: ad ogni passo viene scelta la variabile e il valore di taglio che fornisce la miglior separazione tra segnale e fondo. Si continuano a costruire dei nodi finché non si raggiunge un minimo numero di eventi (dettato da criteri statistici).

3.2.2 Overtraining

Si è detto come la procedura di *training* porti a definire dei parametri che meglio assicurano la distinzione tra segnale e fondo. Sorgono tuttavia dei problemi quando la situazione presenta troppo pochi gradi di libertà: cioè quando troppi parametri vengono impostati con pochi punti (cioè con pochi eventi di *training*).

Se ciò accade, il metodo sembra avere un irragionevolmente ottima performance se testato sui dati usati anche per il *training* e dimostra un comportamento molto peggiore se testato con dati indipendenti. Questo fenomeno è detto *overtraining* cioè sovra-allenamento: il nome deriva dal fatto che alcuni metodi (come le reti neurali o i *decision trees*) possono essere allenati all'infinito, con gli stessi eventi, ed occorre un criterio di stop per il *training*. L'*overtraining* tuttavia è un concetto più generale, deriva dal suddetto problema dei gradi di libertà e può essere riscontrato in qualsiasi metodo di selezione.

L'*overtraining* è comunque fortemente dipendente dal metodo in questione: se da un lato il discriminante di Fisher ne è praticamente *immune* (avendo pochi parametri aggiustabili) i *decision trees* (avendo questi un numero di parametri che può crescere a piacere) hanno sempre *overtraining* se non si usano accorgimenti per evitarlo.

Un parametro che condiziona l'*overtraining* è la statistica: se vengono usati pochi eventi per settare un metodo, questo soffrirà quasi certamente di *overtraining*, mentre con alte statistiche questo è destinato a scomparire. Questo è ovviamente dovuto al fatto che pochi eventi non riusciranno a rappresentare le distribuzioni delle variabili che si avranno in generale in grandi popolazioni, e porteranno il metodo ad essere troppo specializzato verso quei pochi eventi.

L'alta statistica è importante anche per un altro motivo: per individuare l'*overtraining* è necessario avere a disposizione almeno due campioni di dati indipendenti, ma possibilmente ne occorrono tre: uno per effettuare il *training*, uno per individuare quando interromperlo (nei casi in cui questo occorre) e uno per testare le performance.

Vi è da sottolineare che quando si scopre che il proprio metodo soffre di *overtraining* non è detto che lo si debba abbandonare: tuttavia nel valutare i propri risultati occorrerà tenere conto della sovrastima nelle performance, e possibilmente cercare di ottenere più statistica per risolvere il problema.

3.2.3 Ottimizzazione e Criterio di Punzi

L'ottimizzazione dei criteri di selezione è un problema complesso, in particolare se applicato alla selezione di eventi rari. Molti criteri vengono spesso usati per scegliere il livello migliore di selezione, in generale si hanno i seguenti estimatori:

Efficienza (di selezione): definita come il rapporto di eventi di segnale selezionati fratto tutti gli eventi di segnale

$$\varepsilon = \frac{N_{sel}^{sig}}{N_{tot}^{sig}} \quad ; \quad (3.27)$$

Purezza: definita come la frazione di eventi di segnale selezionati, rispetto a tutti gli eventi selezionati

$$p = \frac{N_{sel}^{sig}}{N_{sel}^{sig} + N_{sel}^{back}} \quad ; \quad (3.28)$$

Significanza: definita come la frazione di eventi di segnale selezionati, rispetto alla radice di tutti gli eventi selezionati

$$s = \frac{N_{sel}^{sig}}{\sqrt{N_{sel}^{sig} + N_{sel}^{back}}} \quad . \quad (3.29)$$

Ovviamente, in maniera ideale, si vorrebbe una selezione il più efficiente e pura possibile, tuttavia le due caratteristiche spesso sono in competizione nella realtà per cui bisogna effettuare una scelta. La scelta dipende dal tipo di problema: se si vogliono misurare delle caratteristiche di un evento occorre ottenere più eventi possibili e quindi si sceglierà una selezione meno pura ma più efficiente; viceversa nel caso di decadimenti rari è necessario essere il più possibile liberi da eventi di fondo, perché questi portano errori

nel determinare l'occorrenza degli eventi di segnale, quindi si sceglierà una selezione più pura e meno efficiente.

Ma anche questo passo non è privo di problemi, si può dimostrare che la purezza e la significanza, spesso usati, non sono i criteri migliori per scegliere una selezione a basso fondo. Punzi [33] ha dimostrato che un criterio migliore è la scelta della selezione che fornisce la miglior stima della sezione d'urto di un decadimento misurato, o il miglior limite nel caso non venga trovato segnale.

Data una certa categoria di eventi di segnale, consideriamo due teorie: H^0 che non prevede l'esistenza del segnale e H^m quella che lo prevede, esse prediranno la selezione di un numero medio di eventi (solo fondo in un caso, fondo più segnale nell'altro). Le distribuzioni del numero di eventi saranno poissoniane, approssimabili con gaussiane. Esisterà un numero minimo di eventi n_{min} oltre il quale si parlerà di presenza di segnale. Essendo le due distribuzioni (verosimilmente) in parte sovrapposte, entrambe le teorie prediranno una probabilità, diversa, per l'occorrenza di n_{min} eventi; in particolare n_{min} disterà a e b deviazioni standard dalla media di H^0 e H^m rispettivamente. Si può mostrare che la sezione d'urto misurabile più piccola, cioè la migliore sensibilità o il miglior limite si hanno massimizzando la seguente espressione:

$$\Pi = \frac{\varepsilon(t)}{b^2 + 2a\sqrt{B(t)} + b\sqrt{b^2 + 4a\sqrt{B(t)} + 4B(t)}} \quad (3.30)$$

dove con (t) sono indicate tutte le funzioni dipendenti dalla selezione, ε è l'efficienza e B il numero di eventi selezionati. Questa espressione può essere semplificata se si sceglie $a = b$, diventando:

$$\Pi = \frac{\varepsilon(t)}{a/2 + \sqrt{B(t)}} \quad (3.31)$$

essa è stata usata come criterio per scegliere la selezione migliore nel nostro caso.

3.3 Implementazione e calcolo

Lo studio dei metodi sopra descritti è stato effettuato sotto il *framework* ROOT, descritto nel §3.3.3. Alcuni metodi sono stati implementati diret-

tamente, per uno studio approfondito, per altri si è usato il programma TMVA che verrà descritto approfonditamente nel §3.3.4 e che li conteneva già implementati per il lavoro con ROOT.

3.3.1 Implementazione

Verrà descritto di seguito il modo in cui si implementano dei metodi di questo tipo prendendo ad esempio il metodo PCA del §3.1.3, ma tenendo conto che a parte casi particolari, l'implementazione è sempre simile.

Innanzitutto se è necessaria una trasformazione delle variabili di input, come nel caso della *gaussianizzazione* nel metodo PCA, questa è la prima a dover essere attuata. L'implementazione è banale: si tratta di scrivere la trasformazione e applicarla ad ogni variabile di input, sia per il segnale che per il fondo. Nel caso di decorrelazione delle variabili il metodo è descritto in §3.1.2, e non richiede nessun particolare metodo in implementazione.

Successivamente, il punto di partenza dei vari metodi è la costruzione delle funzioni di distribuzione di probabilità (PDF): avendo a disposizione un numero limitato di dati, e spesso distribuzioni dalle forme non analitiche, la via del fit è impensabile, comunemente si ricorre dunque agli istogrammi, facilmente implementabili nei programmi.

Il problema degli istogrammi è spesso la loro natura *a gradini*, esistono però delle tecniche a cui si può ricorrere per renderli più vicini a distribuzioni analitiche: spesso si ricorre a delle interpolazioni per mezzo di funzioni *spline* (polinomi di alto grado): questi rendono l'istogramma più liscio, e quindi più verosimile, al prezzo del pericolo di rimuovere dei picchi stretti. L'interpolazione per mezzo di *spline* è da utilizzare dunque quando questo pericolo è noto non sussistere.

Tramite le PDF così calcolate si possono costruire le variabili associate ai metodi basati sulle probabilità (PCA, Likelihood), oppure si possono effettuare le scelte su alcuni parametri discriminanti (Decision Trees).

I metodi basati sui singoli eventi (come ad esempio le Reti Neurali) non hanno bisogno delle PDF; questi saranno invece basati sulle ottimizzazioni dei parametri associati, che andranno effettuate confrontandosi con gli eventi.

3.3.2 Valutazione dei metodi

La valutazione e il confronto dei metodi sono stati effettuati basandosi su curve che riportassero la reiezione del fondo rispetto all'efficienza sul segnale.

Per valutare questi parametri sono state utilizzate le curve di distribuzione, per segnale e fondo, delle variabili di output dei vari metodi, sulle quali è stata effettuata una scansione in base al valore di taglio.

L'efficienza sul segnale è stata calcolata semplicemente secondo la (3.27).

Il discorso per quanto riguarda la reiezione del fondo è più complesso. Per un confronto efficace con ciò che ci si aspetta nella realtà è stata scelta come variabile il numero di eventi di fondo residui che si avrebbero in un anno di presa dati dopo la selezione. Questo numero va calcolato come il numero di eventi di fondo selezionati nel caso in studio, moltiplicato per il rapporto tra il numero di quelli analizzati e il numero di quelli attesi in un anno di presa dati; in formule:

$$N_{Sel,1y}^{(B)} = N_{Sel}^{(B)} \frac{N_A^{(B)}}{N_{Tot,1y}^{(B)}} \quad (3.32)$$

$s = N_A^{(B)} / N_{Tot,1y}^{(B)}$ verrà detto *fattore di scala*.

A questo punto sorge un problema: tipicamente nelle analisi di questo tipo ci si vuole porre in condizioni di fondo minimo, cosicché il numero di eventi selezionati diventa esiguo; conseguentemente aumenta l'incertezza sul numero degli eventi selezionati in un anno di presa dati, soprattutto se il fattore di scala è grande.

Per diminuire questa incertezza si è utilizzato un modo per stimare più correttamente il numero di eventi di fondo selezionati. Anziché contare banalmente il numero di eventi che passano la selezione, si è preferito effettuare un *fit* della forma della distribuzione della variabile di output dei vari metodi e calcolare il numero di eventi selezionati come l'integrale degli eventi sottesi dalla curva oltre il valore di taglio. Al fine di effettuare questo fit nella maniera più semplice, ma anche più vantaggiosa, come già detto in §3.1.1 si è preferito trasformare la variabile di output di alcuni metodi.

Una volta ottenute le curve del numero di eventi di fondo residui in funzio-

ne dell'efficienza sul segnale, si è potuto confrontare l'andamento per scegliere i metodi migliori, cioè quelli con meno eventi a parità di efficienza.

Oltre a questo è stato effettuato uno studio per i vari metodi sui limiti ponibili sul segnale, valutando per ciascun metodo il miglior limite ottenibile con il criterio di Punzi.

3.3.3 ROOT

Come si è detto, per l'implementazione computazionale si è usato il software ROOT[34]: esso è un *framework* basato su una programmazione orientata agli oggetti (OOP) in linguaggio C++.

Comprende vari sottoinsiemi di lavoro per la simulazione, la ricostruzione e l'analisi di dati. Comprende un insieme di classi strutturate e specializzate per la Fisica delle alte energie, ma non solo.

Nel particolare è dotato di un'interprete (CINT) del linguaggio C++, che può essere usato in maniera interattiva per brevi macro. Inoltre le classi possono essere usate per costruire dei normali eseguibili.

ROOT può essere usato per molti differenti scopi dall'analisi alla simulazione dati, si rimanda alla referenza citata per ulteriori informazioni.

Per questo lavoro si è utilizzata la versione ROOT v. 5.14. In particolare si sono scritti dei codici autonomi che utilizzassero le classi di ROOT, questo perché i programmi così compilati, oltre che impiegare tempi macchina notevolmente più brevi, sono molto più stabili e sicuri rispetto al modo interpretato che è stato utilizzato solo per brevi macro di test.

3.3.4 TMVA

TMVA (o *Toolkit for MultiVariate Analysis*)[35, 36] è un ambiente, già integrato con ROOT per la valutazione e l'utilizzo di numerosi metodi di analisi multivariata.

Il pacchetto attuale include i seguenti metodi:

- ottimizzazione di tagli rettangolari;
- Likelihood;

- analisi a discriminanti lineari (Fisher e H-Matrix);
- reti neurali artificiali;
- Decision Trees.

Alcuni di questi non sono stati utilizzati nel presente lavoro: in particolare non si sono usati i “tagli rettangolari ottimizzati”, perché questo tipo di metodo è stato ampiamente testato e utilizzato e perché è con questo che si è effettuato il confronto degli altri metodi multidimensionali.

Il pacchetto TMVA consiste in un’implementazione dei metodi suddetti e fornisce la possibilità di effettuare il *training*, di testarli, di applicarli e anche alcune semplici macro di visualizzazione.

TMVA: *training* e test

Il *training* all’interno di TMVA viene effettuato a partire da dei campioni di composizione nota: di questi si selezionano delle variabili che vengono fornite al programma sotto forma o di file di testo o di *n-tuple* per ROOT.

Le variabili specificate possono anche essere fornite sotto forma di espressioni (funzioni, combinazioni) delle variabili di input contenute nei file. Questa trasformazione delle variabili viene effettuata *online*, cioè evento per evento sui valori delle variabili utilizzate. Contestualmente può essere applicata anche la decorrelazione lineare delle variabili descritta nel §3.1.2. Queste medesime trasformazioni verranno attuate, ovviamente anche nella successiva fase di applicazione.

Possono essere specificati dei pesi generali per i dati di segnale e fondo, o anche dei pesi specifici (evento per evento).

I campioni di eventi di input, sia di segnale che di fondo, vengono divisi in due sotto-campioni indipendenti che verranno usati per effettuare il *training* ed il test.

Si può quindi specificare i metodi di cui sopra, ciascuno con varie opzioni personalizzabili.

Viene effettuato il *training* tramite gli eventi dell’apposito campione, e viene calcolato l’output specifico y_{meth} . Viene inoltre calcolata una distribu-

zione di probabilità per il segnale e il fondo \hat{y}_S e \hat{y}_B ; con queste viene calcolata la probabilità per l'evento i -esimo di essere segnale secondo la formula:

$$P_{TMVA}(i) = \frac{f_S \hat{y}_S(i)}{f_S \hat{y}_S(i) + (1 - f_S) \hat{y}_B(i)}$$

dove $f_S = N_S / (N_S + N_B)$ è la frazione di eventi di segnale.

I metodi vengono quindi applicati al campione di test e per ognuno vengono calcolate variabili statistiche di valutazione:

- Efficienza sul segnale a tre diverse efficienze sul fondo;
- *separazione*, cioè

$$\frac{1}{2} \int \frac{(\hat{y}_S(x) - \hat{y}_B(x))^2}{\hat{y}_S(x) + \hat{y}_B(x)} dx \quad ;$$

- la significanza della separazione: definita come la differenza tra le medie degli output di segnale e fondo divisa per la somma delle varianze:

$$\frac{\bar{y}_{meth}(S) - \bar{y}_{meth}(B)}{\sigma_S^2 + \sigma_B^2} \quad .$$

Durante questa fase viene effettuato anche un controllo sull'ammontare dell'*overtraining*, descritto nel §3.2.2.

Tutti i parametri dei metodi, calcolati in questa fase di *training*, vengono salvati in appositi file da utilizzare nella fase di applicazione.

TMVA: Applicazione

La fase di applicazione del TMVA consiste nell'utilizzo dei metodi preparati sugli eventi *reali* per i quali i metodi sono stati pensati.

Viene applicata la trasformazione delle variabili e la decorrelazione in base alle richieste effettuate nella fase di *training*.

I risultati ottenuti come output y_{meth} dai singoli metodi, possono essere utilizzati così per effettuare le selezioni.

Nel nostro caso questa fase di applicazione è stata utilizzata, inoltre, per effettuare un ulteriore studio dei metodi, da paragonare inoltre a quelli implementati personalmente, esternamente a TMVA. I metodi di valutazione utilizzati sono quelli descritti in §3.3.2.

Capitolo 4

Studio dei decadimenti

$$B_{d,s}^0 \rightarrow e^\pm \mu^\mp$$

Lo studio dei decadimenti $B_{d,s}^0 \rightarrow e^\pm \mu^\mp$ è partito dal lavoro descritto nella referenza [37]. Nella nota citata è stato studiato questo decadimento all'interno dell'esperimento LHCb tramite una selezione basata su tagli rettangolari e una Likelihood finale.

Il mio lavoro si è concentrato quindi su questa analisi tramite i metodi multivariati esposti nel Capitolo 3.

4.1 Segnale e Fondi

Il segnale, in questo studio, è costituito dai decadimenti, $B_d \rightarrow e^\pm \mu^\mp$ e $B_s \rightarrow e^\pm \mu^\mp$, tuttavia per semplicità è stato generato solamente questo secondo tipo di decadimento, assumendo che le efficienze di selezione e ricostruzione siano le stesse per i due canali.

Sono stati considerati vari tipi di fondo a questo segnale. Il più importante di questi è costituito dal cosiddetto semileptonico cioè eventi $b\bar{b} \rightarrow l_1^+ l_2^-$ ($l_{1,2} = e, \mu$) in cui entrambi gli adroni dotati di quark b decadono semileptonicamente. Questo campione inclusivo di eventi permette di studiare vari tipi di fondi, tenendo conto, contemporaneamente, dei fattori di scala reciproci.

Altri eventi di interesse sono eventi adronici a due corpi (detti appunto

twobody) come $B_{d,s}^0 \rightarrow h_1^+ h_2^-$ e $\Lambda_b \rightarrow h_1^+ h_2^-$, dove $h = \pi, K, p$. Questi decadimenti sono importanti perché hanno la stessa topologia degli eventi di segnale e masse invarianti simili. Ovviamente possono essere rigettati tenendo conto dell'identificazione delle particelle, ma questa può rivelarsi particolarmente problematica, soprattutto per pioni e kaoni che decadono in volo all'interno del rivelatore. Sono stati considerati anche decadimenti come $B^+ \rightarrow J/\psi K^+$ che hanno un grande *branching ratio* e possono essere selezionati come segnale se una particella viene misidentificata, soprattutto considerando il fatto che la risonanza J/ψ decade circa nel 10 % dei casi in due leptoni carichi (e, μ).

Infine si è studiato $B_c \rightarrow J/\psi \mu \nu$, anch'esso possibile fonte di *background* per i motivi sopra esposti.

4.1.1 Simulazione degli eventi

La generazione degli eventi Monte Carlo usati sia per il fondo che per il segnale è la versione DC04 v2r3 [38] (luminosità nominale $\mathcal{L} = 2 \cdot 10^{32} \text{cm}^{-2} \text{s}^{-1}$).

La generazione di questi eventi è stata effettuata tramite i software descritti nel §2.3.10. In particolare sono state usate le versioni GAUSS v15r19, BOOLE v6r5, BRUNEL v24r5, DAVINCI v12r18.¹

Gli eventi sono generati con EvtGen, non sono stati applicati tagli cinematici in produzione tranne che uno in accettazione, cioè i dati sono generati nel cono di LHCb.

La statistica degli eventi studiati è riportata in Tabella 4.1, dove viene riportato, oltre al *branching ratio* del decadimento, il numero degli eventi attesi, entro l'accettazione del rivelatore, in 1fb^{-1} di luminosità integrata nel punto di interazione di LHCb; viene riportato inoltre il numero di eventi generati e la luminosità integrata equivalente.

¹Per la selezione finale degli eventi usati nella nota [37], è stata usata la versione di DAVINCI v12r17; tuttavia le due versioni non hanno differenze dal punto di vista della fisica, e si preferisce usare la v12r18 per questioni di stabilità computazionale.

Tabella 4.1: Statistica degli eventi simulati analizzati

Canale	BR	Eventi attesi in 1 fb^{-1}	Eventi Generati	\mathcal{L} equivalente in (pb^{-1})
B_d^0		$6.8 \cdot 10^{10}$		
B_s^0		$1.73 \cdot 10^{10}$		
$b\bar{b} \rightarrow e^- \mu^+$	$1.1 \cdot 10^{-2}$	$2.4 \cdot 10^9$	5 M	2.1
$b\bar{b} \rightarrow e^+ \mu^-$	$1.1 \cdot 10^{-2}$	$2.4 \cdot 10^9$	5 M	2.1
$B_d^0 \rightarrow \pi^+ \pi^-$	$5.2 \cdot 10^{-6}$	$3.5 \cdot 10^5$	40 k	113
$B_d^0 \rightarrow K^+ \pi^-$	$2 \cdot 10^{-5}$	$1.4 \cdot 10^6$	8 k	5.8
$B_s^0 \rightarrow K^+ K^-$	$2.5 \cdot 10^{-5}$	$4.3 \cdot 10^5$	8 k	18
$\Lambda_b \rightarrow p \pi^-$	$2.1 \cdot 10^{-5}$	$1.3 \cdot 10^6$	8 k	5.9
$\Lambda_b \rightarrow p K^-$	$7.8 \cdot 10^{-5}$	$3.6 \cdot 10^5$	8 k	22
$B_c \rightarrow J/\psi \mu \nu$	$2 \cdot 10^{-5}$	$2 \cdot 10^6$	101 k	48
$B^+ \rightarrow J/\psi K^+$	$1 \cdot 10^{-3}$	$9.7 \cdot 10^4$	200 k	1023
$B_s^0 \rightarrow e^\pm \mu^\mp$			102 k	

Tabella 4.2: Tagli di preselezione applicati ai dati

χ^2 del vertice	< 14
$Angolo(\vec{p}_B, \vec{r}_{BV} - \vec{r}_{PV})$	$< 32 \text{ mrad}$
IP_{B_s}/σ	< 6
$ m(e\mu) - m_{B_s} $	$< 1 \text{ GeV}$

4.2 Preselezione

Prima dello studio della selezione finale, si è applicata sui dati una serie di tagli di preselezione (o *stripping*) che consentisse di avere un campione più maneggevole di dati. I tagli usati sono riportati in Tabella 4.2.

4.3 Selezione degli eventi

La selezione finale degli eventi è stata studiata sulla base dei metodi multi-dimensionali già esposti.

Dopo i tagli della preselezione sono stati analizzati gli eventi per conside-

rare le variabili più adatte alla selezione finale. In particolare si sono studiate varie combinazioni di variabili da usare nei metodi multidimensionali. Per ogni combinazione sono stati usati i metodi di valutazione dei risultati esposti nel §3.3.2.

Il processo di scelta ha rivelato come le performance dei metodi multidimensionali non sempre migliorino con l'aumentare delle variabili a disposizione. Si è notato inoltre come alcune variabili, particolarmente discriminanti non siano completamente sfruttate dai metodi multidimensionali. Questo risultato ci ha portato a considerare l'ipotesi, rivelatasi efficace, di suddividere il processo di selezione finale in due fasi:

1. applicazione di tagli netti su variabili fortemente discriminanti;
2. applicazione dei metodi multidimensionali per sfruttare al meglio l'informazione contenuta in altre variabili.

Questa combinazione ha portato a risultati migliori dei metodi multidimensionali applicati a tutte le variabili.

Dopo questo processo di scelta, le variabili più idonee sono risultate le seguenti:

- $\Delta r = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ tra i leptoni: si tratta di un calcolo della distanza angolare tra i leptoni, definito in maniera relativisticamente invariante, dove ϕ è l'angolo azimutale e η la pseudorapidità delle particelle.
- $p_T(B_s^0)$: momento trasverso (rispetto alla direzione del fascio) del B_s^0 ricostruito.
- *Isolation* del B , definita come

$$I_B = \frac{p_T(B_s^0)}{p_T(B_s^0) + \sum_i (p_T)_i}$$

dove l'indice i corre sulle tracce dell'evento che hanno la significanza della distanza di minimo approccio con i leptoni minore di 2. Questa variabile tiene conto di quanto una particella è isolata valutando il rapporto dei momenti trasversi. In eventi di segnale *puliti*, chiaramente,

essa dev'essere vicina a 1, mentre in eventi *sporchi*, come quelli adronici, la particella ricostruita sarà contornata da altre di p_T simile.

- Significanza della distanza del vertice del B dal vertice primario, in formule: $(\vec{r}_B - \vec{r}_{PV})/\sigma$, dove σ è l'errore. Questa variabile serve ad escludere gli eventi con particelle provenienti dal vertice primario, anziché dal B ; viene usata la significanza, per tenere contemporaneamente conto della bontà della ricostruzione.
- Significanza del parametro d'impatto del B rispetto al vertice primario, in formule: IP_B/σ .
- $DLL(\mu - \pi)$, ossia *Delta-Log-Likelihood* delle variabili di identificazione delle particelle, in particolare di discriminazione tra muoni e pioni.
- $DLL(e - \pi)$, stessa variabile di sopra ma riferita agli elettroni.
- Angolo di *Pointing*, ossia angolo tra il vettore di impulso del B e la congiungente tra il vertice primario e il vertice del B , in formule:

$$\text{Angolo}(\vec{p}_B, \vec{r}_{BV} - \vec{r}_{PV}) \quad . \quad (4.1)$$

- χ^2 del fit del vertice del B .

Le distribuzioni di alcune di queste variabili sono mostrate nelle Figure 4.1 e 4.2.

Come già detto alcune variabili sono state utilizzate come tagli separati; questo è stato fatto per il χ^2 del vertice, l'angolo di *Pointing* e le DLL . I tagli applicati a queste variabili sono riassunti in Tabella 4.3. Le DLL sono state escluse dai metodi multidimensionali anche per un altro motivo: si è preferito avere sotto controllo questa variabile nell'analisi finale per valutare l'influenza sulla costruzione dei metodi multidimensionali, si affronterà una discussione su questo aspetto nel §4.4.

Le rimanenti variabili citate sono state inserite nei metodi multidimensionali.

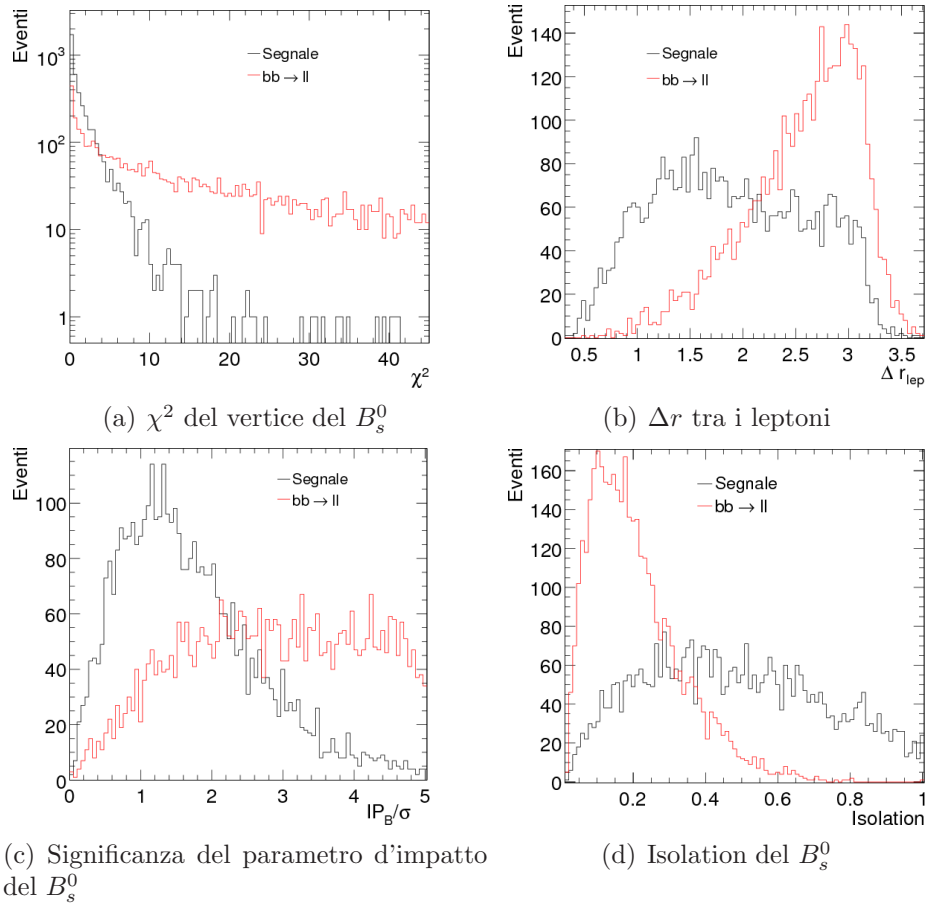


Figura 4.1: Distribuzioni delle variabili utilizzate nell'analisi, per segnale (in nero) e fondo (in rosso). Ogni grafico contiene lo stesso numero di eventi per entrambi i campioni.

Tabella 4.3: Tagli applicati ai dati nella selezione finale

$Angolo(\vec{p}_B, \vec{r}_{BV} - \vec{r}_{PV})$	< 10 mrad
χ^2 del vertice	< 10
$DLL(\mu - \pi)$	> -10
$DLL(\mu - e)$	> -2

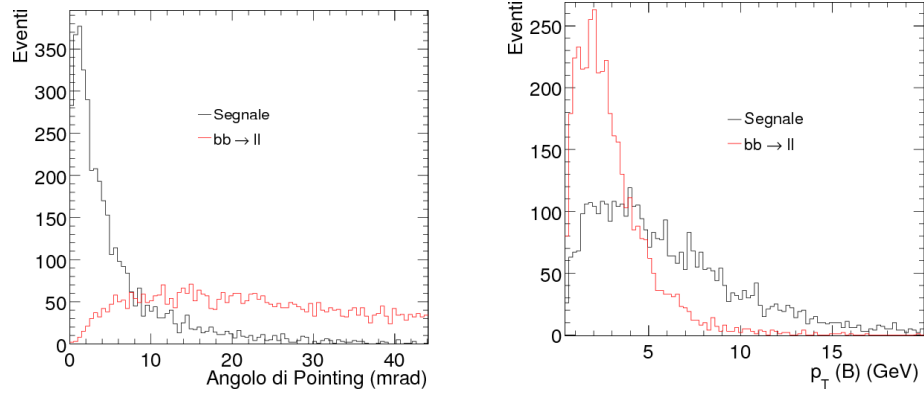
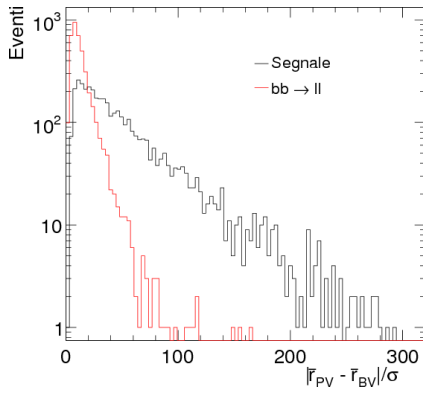
(a) $\text{Angolo}(\vec{p}_B, \vec{r}_{BV} - \vec{r}_{PV})$ (di pointing)(b) Impulso trasverso del B_s^0 (c) Significanza della distanza tra vertice primario e vertice del B

Figura 4.2: Distribuzioni delle variabili utilizzate nell'analisi, per segnale (in nero) e fondo (in rosso). Ogni grafico contiene lo stesso numero di eventi per entrambi i campioni.

Tabella 4.4: Stima dell'*overtraining* per i metodi testati come differenza tra l'efficienza sul segnale ottenuta con il campione di training e quella ottenuta sul campione di test

Metodo	Overtraining
Likelihood	0.002
Likelihood Decorrelata	-0.001
Discriminante di Fisher	0.008
Rete Neurale	0.003
Decision Trees	0.089
PCA	0.003

4.3.1 Applicazione dei metodi multidimensionali

I metodi multidimensionali descritti nel Capitolo 3 sono stati preparati e applicati a questo decadimento.

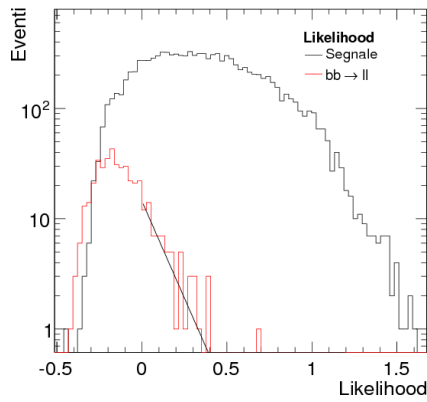
Il canale di fondo usato per il *training* delle variabili è stato il semileptonico inclusivo. Questo perché i fondi adronici costituivano una fonte trascurabile di fondo e contaminavano la costruzione dei metodi multidimensionali, si riparerà di questo nel §4.4.

I risultati di alcune variabili di *output* sono mostrati in Figura 4.3, dove si possono notare anche i fit applicati alle distribuzioni delle variabili di fondo.

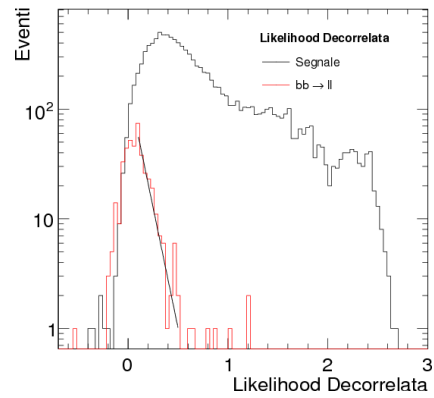
Sono stati calcolati quindi i valori di efficienza sul segnale e reiezione del fondo per valutare i metodi; i risultati sono esposti in Figura 4.3.1. Si può notare come il metodo migliore sia risultata la Likelihood a variabili decorrelate (o Likelihood Decorrelata).

Inoltre si è data una valutazione dell'*overtraining* (Cfr. §3.2.2) come differenza tra l'efficienza ottenuta con i campioni di training e quella con i campioni di test. I risultati sono mostrati in Tabella 4.4. Si può notare che, come previsto, i *decision trees* sono affetti da un grosso *overtraining*.

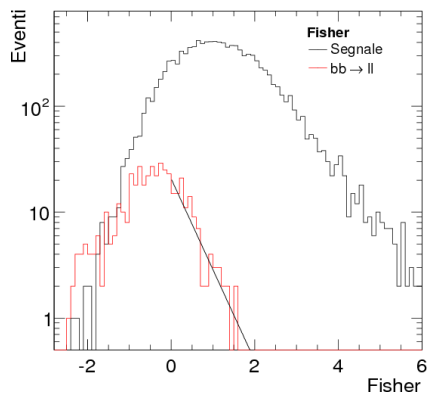
Il taglio finale da applicare alla Likelihood è stato scelto tramite il metodo di Punzi (Cfr. §3.2.3), secondo le curve di distribuzione per il parametro dell'eq. (3.31) in funzione dell'efficienza sul segnale (come mostrato in Figura 4.5).



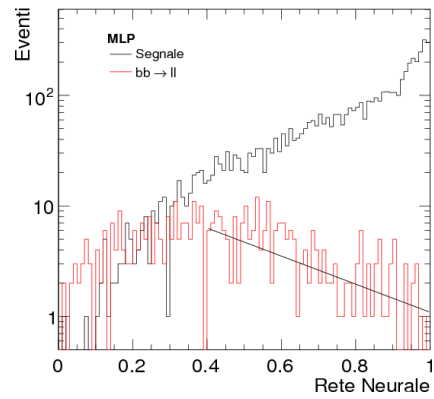
(a) Likelihood



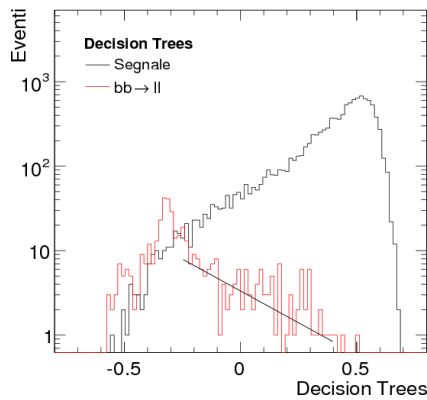
(b) Likelihood a variabili decorrelate



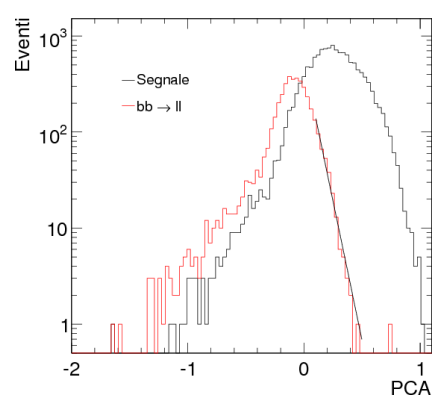
(c) Discriminante di Fisher



(d) Rete neurale artificiale



(e) Decision Trees



(f) Approssimazione di proiezione e correlazione (PCA)

Figura 4.3: Distribuzione delle variabili di output per i metodi studiati per segnale (istogramma nero) e fondo (istogramma rosso). È inoltre mostrato il fit della distribuzione del fondo (linea nera).

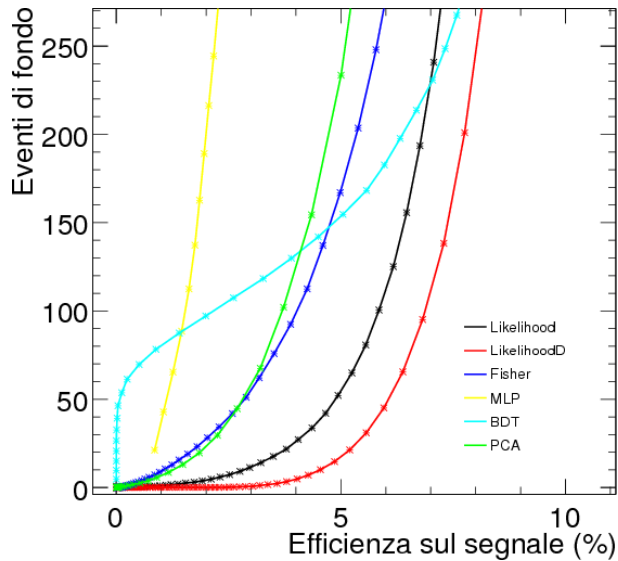


Figura 4.4: Andamento del numero di eventi di fondo selezionati in 1 fb^{-1} di luminosità integrata rispetto all'efficienza sul segnale (in percentuale), per diversi metodi utilizzati.

Infine tramite uno stimatore bayesiano si è calcolato il limite superiore ponibile dall'esperimento al *branching ratio* del decadimento. Il metodo seguito è il seguente: il programma calcola un limite su un processo poissoniano, come quello studiato, con fondo, incorporando gli effetti di incertezza dovuti all'accettazione sul fondo e sul segnale. Viene assunta una *prior* PDF piatta per il rate del segnale e delle *prior* gaussiane per l'accettazione e il fondo atteso [39].

4.4 Strategia di applicazione dell'analisi ai dati reali

L'applicazione ai dati reali di un'analisi di questo tipo risulta non banale. È da tenere in considerazione infatti la problematica del *training*. Per essere il più possibile efficaci nel settare i metodi multidimensionali, infatti, bisogna effettuare il *training* con i dati reali: tuttavia nella realtà non si conosce quali dati appartengono al segnale e quali al fondo.

Nella pratica della fisica delle alte energie solitamente si studia la com-

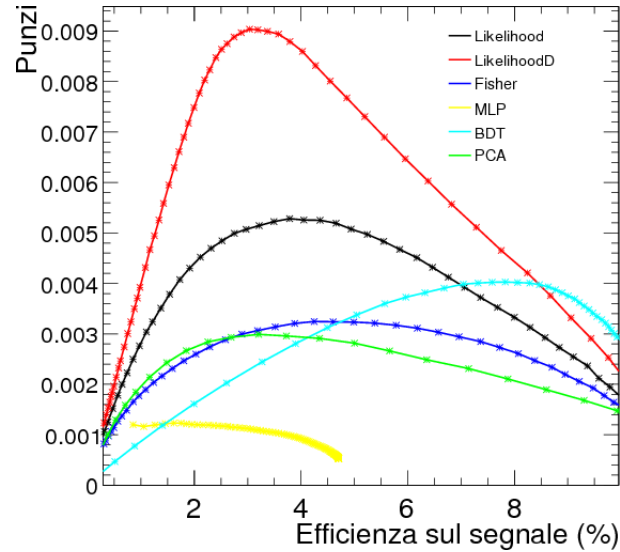


Figura 4.5: Andamento della variabile associata al criterio di Punzi in funzione dell'efficienza sul segnale. Si può notare per ogni metodo la presenza di un punto di massimo che sarà quello in cui effettuare l'analisi con la maggiore sensibilità.

posizione degli eventi a partire dalla massa invariante delle particelle selezionate. In particolare il segnale ovviamente deve trovarsi nella zona attorno alla massa del B : questa zona è definita solitamente come una finestra più o meno grande a seconda della risoluzione. All'interno della finestra in massa, oltre al segnale e a seconda della selezione, si trova una quantità più o meno grande di fondo che dev'essere stimata; la stima viene effettuata attraverso le cosiddette *sidebands*: queste sono le zone di massa invariante immediatamente vicine a quella selezionata. In genere il fondo ha una distribuzione lievemente esponenziale che diminuisce verso le alte masse ed è rappresentato efficacemente dalle *sidebands*. Questo tuttavia non succede se anche nel fondo sono presenti dei picchi in massa invariante, essi distorcono la selezione rendendo più difficile rappresentare le distribuzioni del fondo nella finestra di massa.

Nel nostro caso la questione è da separare in due parti: per quanto riguarda il fondo semileptonico inclusivo, dominante, la distribuzione delle variabili è circa uniforme tra *sidebands* e finestra di massa e quindi non causa problemi.

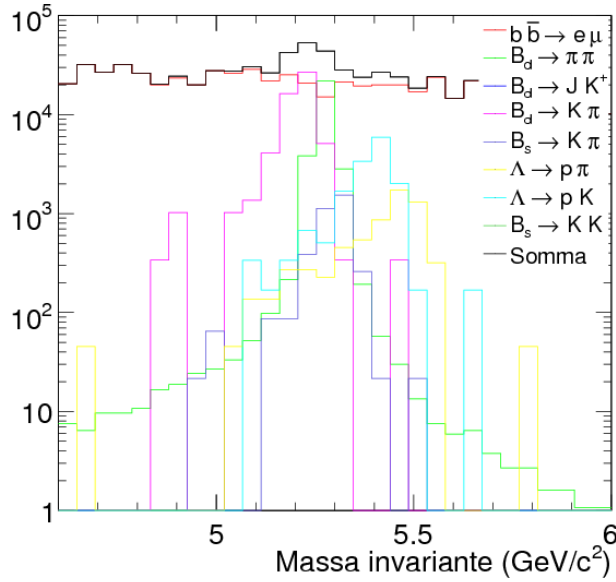


Figura 4.6: Distribuzione di massa invariante per i diversi fondi del decadimento studiato. Si può notare che mentre la distribuzione del semileptonico inclusivo (in rosso) è uniforme, i fondi *two body* presentano dei picchi che possono rendere difficoltosa la stima del fondo a partire dalle *sidebands*.

Diverso è il discorso quando si parla dei fondi *two body*, essi, come si può notare in Figura 4.6, mostrano dei picchi all'interno della massa invariante. Si è quindi deciso di studiare le potenzialità di questa analisi al variare della presenza nel fondo di frazioni maggiori di *two body* (chiameremo questa presenza *contaminazione*, perché contamina la forma della likelihood ottenuta assumendo che il fondo sia esclusivamente semileptonico).

Ovviamente il livello di contaminazione dipende dal taglio in DLL delle due particelle che si applica ai dati. Più stretto è il taglio, minore la contaminazione. In Tabella 4.5 sono mostrati alcuni livelli significativi di contaminazione (definiti come percentuale di *two body* rispetto al fondo complessivo) e i valori del taglio in DLL corrispondenti.

Tramite questo tipo di studio si è visto che è possibile ridurre il livello di fondo *two body* ad un livello trascurabile, almeno dal punto di vista delle distribuzioni delle variabili da utilizzare nei metodi multidimensionali.

Fatto questo sorge un problema analogo per il segnale: come costruire le distribuzioni *reali* per il segnale? L'unica possibilità per questo tipo di deca-

Tabella 4.5: Tagli applicati alle DLL dei leptoni e livelli di contaminazione relativa del fondo *two body* rispetto al leptonico

Tagli in DLL		Contaminazione
$DLL(\mu - \pi) > -10$	$DLL(e - \pi) > -2$	4 %
$DLL(\mu - \pi) > 0.1$	$DLL(e - \pi) > 1.5$	1 %
$DLL(\mu - \pi) > 2$	$DLL(e - \pi) > 4$	0.6 %

dimento (e in generale per i decadimenti rari) è usare le simulazioni Monte Carlo. Tuttavia se per gli studi preliminari ci si può accontentare di approssimare il segnale, per lo studio degli eventi reali occorre che gli eventi Monte Carlo riproducano efficacemente il segnale. L'unico modo per valutare la bontà delle simulazioni è usare dei cosiddetti *canali di controllo*: ossia altri decadimenti facilmente studiabili nei dati reali che possano rendere conto della variazione tra simulazioni e realtà e delle possibili distorsioni di accettazione e ricostruzione dell'esperimento. I canali di controllo devono essere, ovviamente, più simili possibile al segnale dal punto di vista topologico.

Nel nostro caso la scelta ricade sui seguenti canali: $B_s^0 \rightarrow \pi^+\pi^-$, $B^+ \rightarrow J/\psi(\rightarrow e^+e^-)K^+$ e $B^+ \rightarrow J/\psi(\rightarrow \mu^+\mu^-)K^+$. Le distribuzioni delle variabili cinematiche e topologiche di questi canali risultano infatti molto simili al segnale.

La procedura per capire quindi l'andamento del segnale nei dati reali può essere eseguita in due modi: o una ricalibrazione del generatore degli eventi Monte Carlo (un approccio per così dire *a monte*, laborioso ma di sicura efficacia), o una ricalibrazione delle distribuzioni delle variabili già ottenute (approccio *a valle*, più semplice ma anche meno sicuro). Su questo aspetto dell'analisi vi è ancora del lavoro da fare, da parte della collaborazione LHCb, per la decisione della strategia più adatta.

4.5 Risultati

I decadimenti $B_{d,s}^0 \rightarrow e^\pm\mu^\mp$ studiati, sono stati selezionati con una combinazione di tagli rettangolari e metodi multidimensionali. In particolare si è

trovato come metodo migliore, e applicato, la Likelihood Decorrelata.

Tramite questa selezione è possibile stabilire il limite che l'esperimento LHCb sarà in grado di porre al *branching ratio* di questo decadimento.

Il taglio sulla likelihood è stato ottimizzato tramite il metodo di Punzi; il limite sul *branching ratio* al 90% di Confidence Level è pari a $1.1 \cdot 10^{-8}$ per $B_s^0 \rightarrow e^\pm \mu^\mp$ e di $2.7 \cdot 10^{-9}$ per $B_d^0 \rightarrow e^\pm \mu^\mp$.

Capitolo 5

Studio del decadimento

$$B_s^0 \rightarrow \mu^+ \mu^- \gamma$$

A seguito di quanto si è imparato nello studio del decadimento $B_s^0 \rightarrow e^\pm \mu^\mp$ si è deciso di effettuare un'analisi multidimensionale anche per il decadimento $B_s^0 \rightarrow \mu^+ \mu^- \gamma$.

5.1 Segnale e fondo

Il segnale studiato è costituito da $B_s^0 \rightarrow \mu^+ \mu^- \gamma$, tuttavia la stessa selezione può essere pensata anche per il B_d^0 . Infatti l'efficienza di selezione sul segnale e la reiezione del fondo dovrebbero essere le stesse.

La rivelazione dei muoni è semplice all'interno dell'esperimento LHCb: generalmente le tracce sono completamente ricostruite e l'identificazione come particelle è buona. Con i due muoni si può quindi ricostruire subito il vertice del B_s^0 tramite fit.

Qualche problema può sorgere nella ricostruzione del fotone, e quindi del decadimento completo. Innanzitutto per il fotone non si ha una vera e propria traccia ma solo la sua deposizione di energia nel calorimetro elettromagnetico. Questa informazione è costituita essenzialmente, oltre che dall'energia rilasciata, da un punto nello spazio, ma non da una direzione. Per questo motivo la ricostruzione del fotone porta ad associargli un impulso diverso a seconda del punto da cui esso si ritiene che provenga. Nella fase di ricostruzione, in prima istanza i fotoni vengono associati al vertice primario; di

conseguenza per ricostruire il decadimento bisogna innanzitutto ricalcolare l'impulso dei fotoni facendo sì che essi provengano dal vertice dei due muoni. Questo da un lato può portare ad un peggioramento nella risoluzione sulla massa e sull'impulso del B_s^0 , venendo ad essere dipendenti dal fit del vertice dei soli muoni, ma dall'altro porta a rigettare più facilmente le combinazioni spurie che portano a B_s^0 non provenienti dal vertice primario, e quindi rigettabili, come vedremo, con un taglio in angolo di puntamento.

Come fondo è stato studiato il canale inclusivo $b\bar{b} \rightarrow \mu^+ \mu^-$ che comprende tutti gli eventi in cui nello stato iniziale è presente una coppia di quark b e nello stato finale i due muoni. Il campione comprende sia gli eventi in cui i muoni provengono da decadimenti dei B ma anche quelli in cui essi derivano da altre particelle provenienti dal vertice primario.

Il fondo allora sarà costituito essenzialmente da combinazioni casuali di questi due muoni con fotoni presenti nel rivelatore e provenienti o dal vertice primario o dal decadimento di pioni neutri presenti nello stesso evento.

Statisticamente il fondo studiato rappresenta il 2.6 % del canale $b\bar{b}$ inclusivo e, grazie alla buona identificazione e ricostruzione dei muoni, dovrebbe essere la sorgente di fondo dominante per questo decadimento.

5.1.1 Simulazione degli eventi

Il segnale per questo studio, non ancora presente all'interno dei dataset standard dell'esperimento LHCb è stato generato appositamente.

La generazione è stata effettuata con il metodo *phase space*. Questo non tiene in conto delle ampiezze di elicità del decadimento specifico, ma lo genera solo considerando lo spazio delle fasi disponibile per le particelle nello stato finale. Questo metodo è stato applicato in quanto non erano disponibili le routine per il programma EvtGen che tenessero conto delle ampiezze di elicità.

Erano tuttavia disponibili alcune simulazioni Monte Carlo del decadimento nel vuoto basate sulle ampiezze di elicità. Queste sono state confrontate con le nostre simulazioni per una valutazione della differenza delle distribuzioni delle variabili.

La differenza più sensibile risiede nella massa invariante dei due muoni: questa presenta i picchi delle risonanze negli eventi simulati con ampiezze di elicità, picchi non presenti negli eventi simulati *phase space* (Figura 5.2).

Le variabili utilizzate nel nostro studio per i metodi multidimensionali (come ad esempio l'energia trasversa del fotone, mostrata in Figura 5.1) non si discostano molto da quelle calcolate con ampiezze di elicità; dunque anche se sarebbe necessario un ripesamento si può considerare, in prima approssimazione, che i risultati di questa analisi non dovrebbero variare sensibilmente.

Per il fondo si sono utilizzati i dati “DC06-phys-v2-lumi2” [38] contenenti una coppia di muoni sempre in accettazione. Il loro metodo di produzione è un pò particolare: per ogni coppia $b\bar{b}$ si procede all'adronizzazione per 10 volte e si accettano gli eventi se contengono due muoni in accettazione. Questo è stato fatto nella produzione degli eventi Monte Carlo per avere un'efficienza maggiore. Dal punto di vista pratico non cambia nulla rispetto a una risimulazione completa di ogni evento, salvo la possibilità che ci siano delle correlazioni tra eventi provenienti dalla stessa coppia di quark. Tuttavia la probabilità che due coppie di muoni provengano dalla stessa coppia con adronizzazioni differenti è praticamente trascurabile. Per una referenza sulla produzione si veda [40] e collegamenti seguenti.

Per la simulazione e ricostruzione completa degli eventi, si sono usate le seguenti versioni dei programmi: GAUSS v25r10, BOOLE v13r3, BRUNEL v31r8 e DAVINCI v19r2, per i quali si rimanda al §2.3.10.

5.2 Preselezione

Per ridurre il numero di eventi di fondo e avere quindi una statistica trattabile si sono attuati alcuni semplici tagli di preselezione. Questi tagli sono riassunti in Tabella 5.1 e spiegati qui di seguito.

Per la preselezione dei muoni si sono considerati quelli che vengono definiti *loose muons*, ossia muoni con basse richieste di identificazione. Si chiede che i

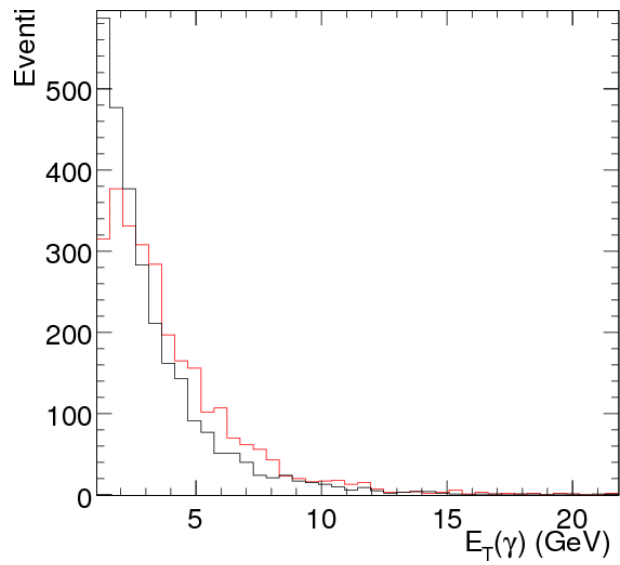


Figura 5.1: Distribuzione dell'energia trasversa dei fotoni nel decadimento $B_s^0 \rightarrow \mu^+ \mu^- \gamma$: eventi simulati tramite ampiezze di elicità (in rosso) ed eventi simulati *phase space* (in nero).

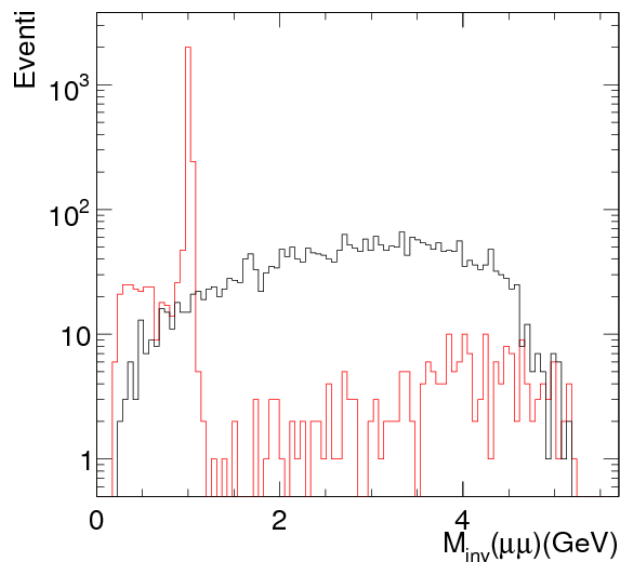


Figura 5.2: Distribuzione della massa invariante dei due muoni nel decadimento $B_s^0 \rightarrow \mu^+ \mu^- \gamma$: eventi simulati tramite ampiezze di elicità (in rosso) ed eventi simulati *phase space* (in nero).

Tabella 5.1: Tagli di preselezione applicati ai dati per la selezione del decadimento $B_s^0 \rightarrow \mu^- \mu^+ \gamma$

p_T dei muoni	> 500 MeV
E_T del fotone	> 1 GeV
$Angolo(\vec{p}_B, \vec{r}_{BV} - \vec{r}_{PV})$	< 14 mrad
$ M_{inv}(\mu\mu\gamma) - M_B $	< 600 MeV

muoni abbiano tracce *lunghe*, ossia tracce che hanno lasciato un segnale lungo tutto il rivelatore. Inoltre si richiede esplicitamente che queste particelle siano state rivelate anche dal *muon system*. Queste richieste sono meno stringenti di un taglio sulla DLL di identificazione, ma sono sufficienti almeno allo stato attuale per questa analisi.

Ai due muoni si è applicato in aggiunta un taglio sul χ^2 del fit del vertice, essenzialmente per rigettare il fondo combinatoriale. Inoltre si è richiesto che essi avessero un impulso trasverso maggiore di 500 MeV. Questa è una semplice richiesta utile principalmente per eliminare gli eventi non interessanti composti da muoni non da mesoni B .

Analogamente si richiede che l'energia trasversa del fotone sia maggiore di 1 GeV. Questa in realtà è una richiesta molto più stringente di quella effettuata sui muoni ed è stata effettuata in virtù del tipo di decadimento studiato. Molti fotoni di basso impulso e bassa energia trasversa provengono da decadimenti $\pi^0 \rightarrow \gamma\gamma$ dei molti pioni presenti nello stesso evento, questi combinandosi con muoni provenienti da una risonanza potrebbero portare a una massa invariante simile a quella di un B . Un taglio in energia porta quindi alle regioni di bassa statistica dei fotoni.

Il taglio in energia trasversa del fotone in realtà potrebbe porre dei limiti alla stima corretta dei contributi del canale $B_s^0 \rightarrow \mu^+ \mu^- \gamma$ al canale senza il fotone. Per far ciò infatti è importante avere delle stime dello spettro di questo decadimento nell'energia del fotone anche per i fotoni meno energetici. Tuttavia per poter rigettare efficacemente il grosso fondo, costituito come detto dai fotoni di decadimento dei pioni, si è dovuto operare in questo modo.

Tabella 5.2: Statistica degli eventi e delle combinazioni di segnale e di fondo analizzati

Segnale				
	Totale	Preselezione	Verità MC	Preselezione e Verità MC
Combinazioni	16455	3509	4518	2815
Eventi	6877	3155	4023	2761
Fondo				
	Totale	Preselezione		
Combinazioni		30235		
Eventi	14907042	26967		

Per gli stessi motivi sopra esposti si è applicato già in preselezione un taglio nell'angolo tra l'impulso del B_s^0 e la congiungente fra il vertice primario e il vertice del B_s^0 . In particolare si è richiesto che quest'angolo, detto come già visto di *pointing*, fosse minore di 14 mrad.

Sempre sull'angolo di *pointing* si è applicata un'ulteriore selezione. Il tipo di decadimento considerato porta ogni evento, anche del segnale, ad essere composto da numerose combinazioni di una o due coppie di muoni con le centinaia di fotoni presenti in un evento tipico. Per questi motivi si è scelto di selezionare una sola combinazione per evento scegliendo quella con il *pointing* migliore, ossia con l'angolo di puntamento minore.

Infine si è richiesto che la massa invariante delle tre particelle non differisse di più di 600 MeV dalla massa del B_s^0 . Questa finestra di massa è stata mantenuta così ampia per una stima corretta del fondo, importante per la costruzione delle variabili multidimensionali. Essa può essere ampiamente ridotta in fase di selezione finale.

Per gli eventi di segnale si è infine richiesto, tramite l'associazione con la verità del Monte Carlo generato, che la combinazione selezionata all'interno di ciascun evento contenesse il fotone esatto del Monte Carlo. Questa è una condizione necessaria per poter costruire dei metodi multidimensionali basati sulle vere distribuzioni del decadimento studiato. Si è notato, tuttavia, che la preselezione di cui sopra non è automaticamente soddisfatta richiedendo

la verità Monte Carlo. Infatti, come si può notare nelle statistiche in Tabella 5.2, gli eventi contemporaneamente preselezionati e Monte Carlo sono meno degli eventi Monte Carlo e degli eventi preselezionati. Questa perdita in efficienza deriva in realtà dalle caratteristiche dell'evento studiato. Infatti la ricostruzione del fotone non consente un'associazione semplice. Le motivazioni sono da ricercarsi in parte nelle performance del calorimetro e in parte nell'incertezza intrinseca dell'associazione di uno sciame con una particella precisa. Questo aspetto dell'analisi dovrà essere approfondito per una valutazione migliore del fondo intrinseco, interno agli eventi segnale.

5.3 Selezione degli eventi

5.3.1 Applicazione dei metodi multidimensionali

La selezione finale degli eventi è stata basata sui metodi multidimensionali studiati precedentemente per il decadimento $B_s^0 \rightarrow e^\pm \mu^\mp$. In particolare si sono applicati a questa analisi i metodi che precedentemente si erano rivelati più performanti.

Si sono studiate le variabili più adatte e si sono scelte quelle che portavano a risultati migliori. In particolare si sono utilizzate:

- impulso trasverso, p_T , del B ;
- significanza del parametro di impatto del B rispetto al vertice primario (IP_{B_s}/σ);
- impulso trasverso del fotone ($p_T(\gamma)$);
- *Isolation* del fotone: calcolata come il rapporto tra l'impulso trasverso del fotone e la somma degli impulsi trasversi delle tracce con $\Delta r < 2$ rispetto al fotone, in formule:

$$I_\gamma = \frac{p_T(\gamma)}{p_T(\gamma) + \sum_i (p_T)_i} .$$

- *Pointing* del B (come definito nella (4.1)).

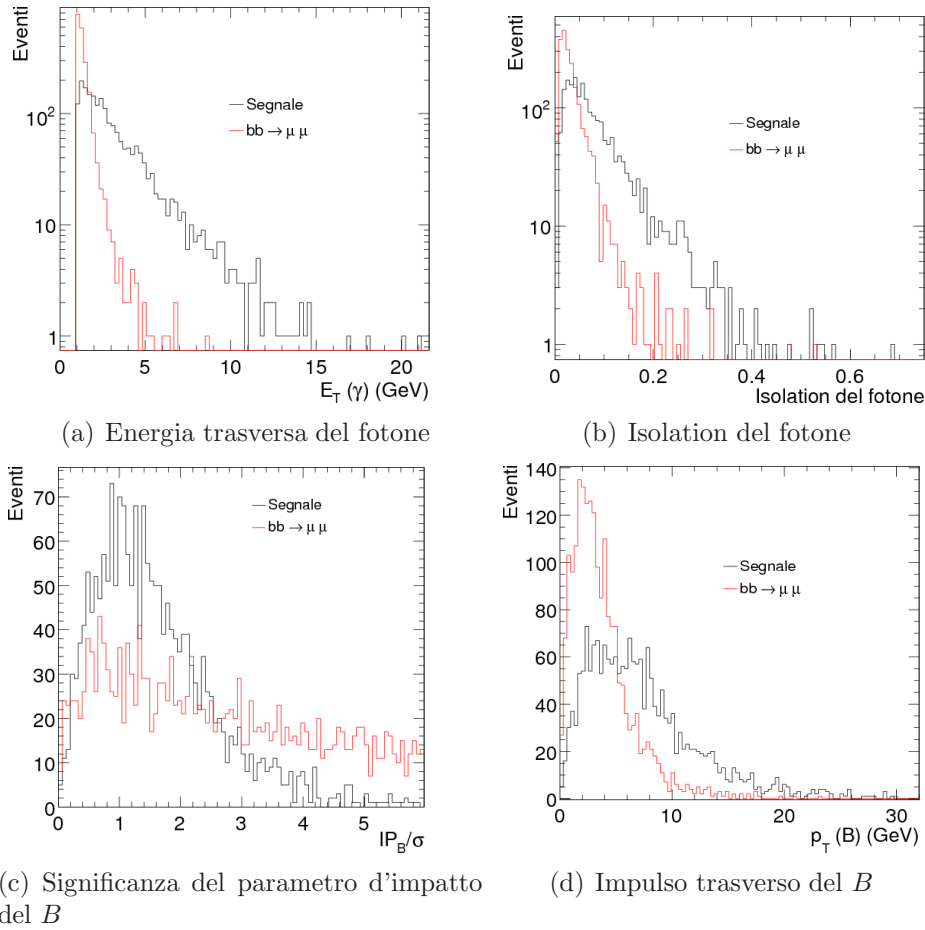


Figura 5.3: Distribuzioni di alcune delle variabili utilizzate nell'analisi del decadimento $B_s^0 \rightarrow \mu^+ \mu^- \gamma$ per segnale (in nero) e fondo (in rosso). Ogni grafico contiene lo stesso numero di eventi per entrambi i campioni.

Si noti che sia il *pointing* che l'impulso trasverso del fotone sono stati inclusi anche nelle variabili da usare nei metodi multidimensionali, nonostante fossero già stati usati come tagli netti per la preselezione. Infatti, visto il loro potere discriminante si è preferito sfruttare tutta l'informazione contenuta in queste variabili.

La procedura seguita per l'applicazione dei metodi multidimensionali è la medesima usata in precedenza per il decadimento $B_s^0 \rightarrow e^\pm \mu^\mp$.

Si è effettuato il training con le variabili scelte e si sono ottenute le distribuzioni delle variabili di output dei vari metodi (mostrate in Figura 5.4).

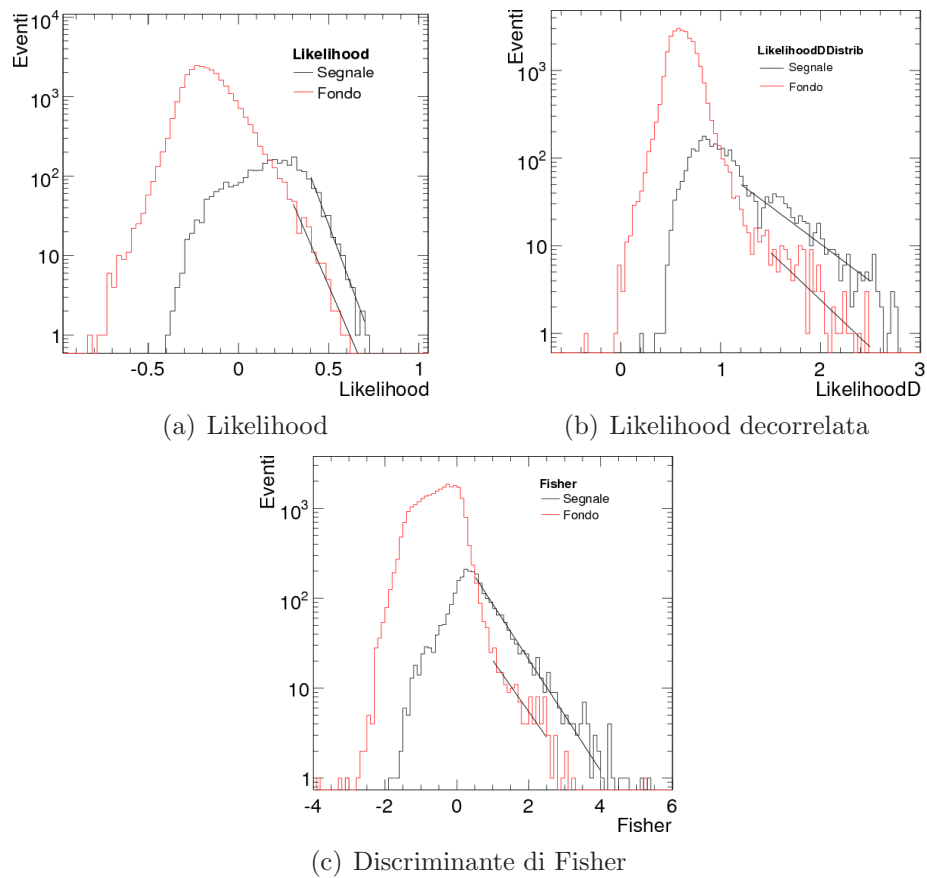


Figura 5.4: Distribuzione delle variabili di output per i metodi studiati per segnale (istogramma nero) e fondo (istogramma rosso). Sono inoltre mostrati i fit delle distribuzioni (linee nere).

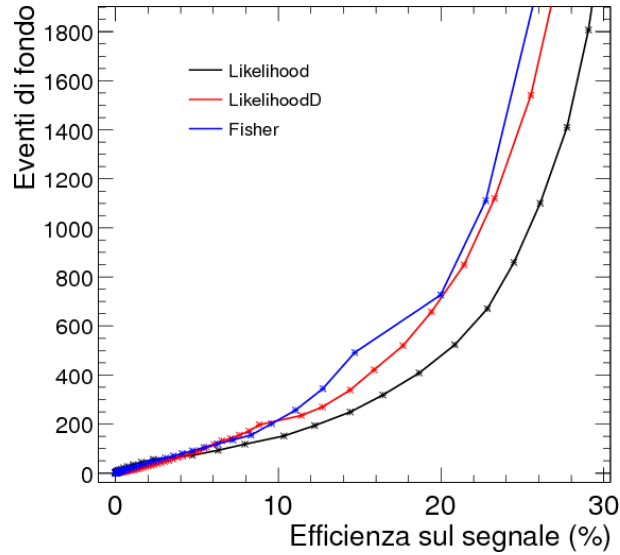


Figura 5.5: Andamento del numero di eventi di fondo selezionati in un anno di presa dati in funzione dell'efficienza sul segnale.

Nel calcolo finale dell'efficienza e della reiezione del fondo è stata apporata una piccola modifica al metodo visto in precedenza. Poiché in questo caso il numero di eventi di fondo disponibili per l'analisi multivariata sopravanzava il numero di quelli di segnale, si è preferito applicare il fit delle variabili di output anche al segnale. Questo per le stesse ragioni per cui era stato precedentemente attuato il fit per il fondo.

Si sono ottenute dunque le curve di distribuzione del numero di eventi di fondo selezionati in un anno di presa dati in funzione dell'efficienza sul segnale (mostrate in Figura 5.3.1). Come si può vedere in questo caso il metodo migliore è risultato la Likelihood a variabili non decorrelate.

Per il calcolo del limite sul *branching ratio* il numero degli eventi di fondo e l'efficienza sul segnale sono state riscalate per tenere conto della finestra di massa: durante la selezione finale si considera una finestra di massa di ± 200 MeV attorno alla massa del B .

Si noti che il numero di eventi mostrato in Figura 5.3.1 tiene conto solo di quelli analizzati e dev'essere pertanto moltiplicato per il fattore di scala corrispondente alla frazione di eventi analizzati sul totale annuo.

Si è quindi calcolato il punto migliore per la stima del limite sul *branching*

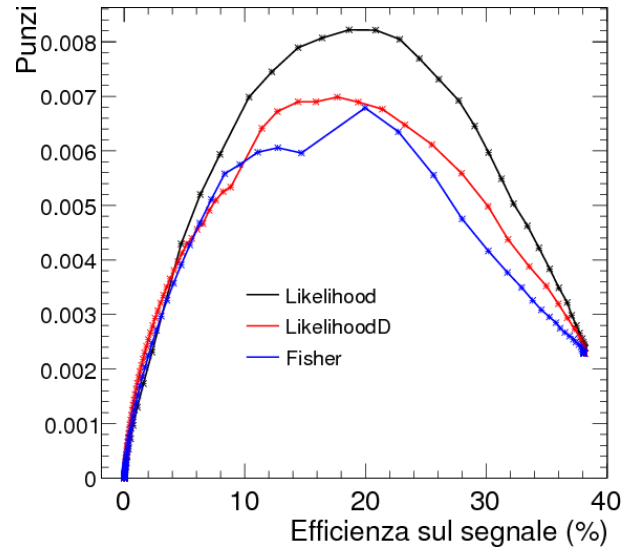


Figura 5.6: Andamento del criterio di Punzi in funzione dell'efficienza sul segnale.

ratio tramite l'andamento del criterio di Punzi. Questo andamento è mostrato in Figura 5.6.

Infine si è calcolato il limite superiore ponibile dall'esperimento in un anno di presa dati al *branching ratio* del segnale studiato, trovando il seguente valore al 90 % di Confidence Level: $BR(B_s^0 \rightarrow \mu^+\mu^-\gamma) < 3.2 \cdot 10^{-7}$; per il canale $B_d^0 \rightarrow \mu^+\mu^-\gamma$ assumendo la stessa efficienza di selezione per il segnale si trova: $BR(B_d^0 \rightarrow \mu^+\mu^-\gamma) < 8.0 \cdot 10^{-8}$.

Nella valutazione dei contributi a corto raggio al decadimento studiato devono essere eliminate le risonanze dei muoni di cui si è parlato nel §5.1.1. Questo porterebbe ad una perdita di efficienza sul segnale; una prima stima indica che l'87% degli eventi preselezionati si trova fuori dalle zone delle risonanze; questa stima dovrà tuttavia essere confortata da un'analisi specifica quando saranno disponibili le simulazioni del decadimento studiato costruite con ampiezze di elicità.

5.4 Risultati

Si sono applicati i metodi multidimensionali studiati ai decadimenti $B_{s,d}^0 \rightarrow \mu^+ \mu^- \gamma$, trovando come metodo migliore la Likelihood.

Tramite il criterio di Punzi si è calcolato a partire dall'efficienza sul segnale e dal numero di eventi di fondo selezionati, il punto migliore di selezione.

Si sono quindi trovati i seguenti limiti, ponibili dall'esperimento LHCb in un anno di presa dati:

$$BR(B_s^0 \rightarrow \mu^+ \mu^- \gamma) < 3.2 \cdot 10^{-7}$$

$$BR(B_d^0 \rightarrow \mu^+ \mu^- \gamma) < 8.0 \cdot 10^{-8}$$

In particolare si può notare come il limite ponibile sul B_d , sia già in un anno migliore rispetto a quello posto dall'esperimento BABAR [14].

Capitolo 6

Conclusioni

Questo lavoro di tesi si è concentrato sullo studio, lo sviluppo e la valutazione di metodi multidimensionali per la selezione di eventi rari, e sulla loro applicazione nell'esperimento LHCb.

In particolare si sono studiati i metodi: Likelihood, discriminante di Fisher, approssimazione di proiezione e correlazione, reti neurali artificiali e *decision trees*. I diversi metodi sono stati studiati tramite l'implementazione di algoritmi in linguaggio C++ basandosi sulle classi del programma ROOT. È stato inoltre studiato e testato il programma di analisi multivariata TMVA, pensato anch'esso per ROOT. Quest'ultimo si è rivelato particolarmente interessante, consentendo un'applicazione sistematica e semplice di numerosi metodi multidimensionali.

I diversi metodi sono stati testati in base a proprietà di efficienza, reiezione del fondo, overtraining e infine in base al criterio di Punzi valutando le condizioni migliori per la misura dei *branching ratios* dei decadimenti.

Si sono studiate le performance dei diversi metodi in relazione a diverse combinazioni di variabili. Si sono ricavate le variabili più adatte ed è stato valutato il metodo migliore in base ai criteri sopra elencati. L'applicazione di una combinazione di tagli netti su variabili molto discriminanti e di metodi multidimensionali su altre variabili si è rivelata la scelta migliore.

Nel particolare i metodi sono stati studiati tramite l'applicazione al decadimento LFV $B_{d,s}^0 \rightarrow e^\pm \mu^\mp$. Il metodo multidimensionale che ha mostrato in questo caso le migliori performance è stata la Likelihood con variabili decor-

relate. Tramite questa selezione è stato possibile calcolare i seguenti limiti superiori ai *branching ratios*, ponibili dall'esperimento LHCb in 1 fb^{-1} di luminosità integrata:

$$\begin{aligned} BR(B_s^0 \rightarrow e^\pm \mu^\mp) &< 1.1 \cdot 10^{-8} \quad ; \\ BR(B_d^0 \rightarrow e^\pm \mu^\mp) &< 2.7 \cdot 10^{-9} \quad . \end{aligned}$$

Quindi i metodi studiati sono stati applicati allo studio dei decadimenti rari $B_{d,s}^0 \rightarrow \mu^+ \mu^- \gamma$. Lo studio ha portato alla scelta della Likelihood come metodo migliore. Si sono calcolati i seguenti limiti ponibili dall'esperimento LHCb sui *branching ratios* di questi decadimenti:

$$\begin{aligned} BR(B_s^0 \rightarrow \mu^+ \mu^- \gamma) &< 3.2 \cdot 10^{-7} \quad ; \\ BR(B_d^0 \rightarrow \mu^+ \mu^- \gamma) &< 8.0 \cdot 10^{-8} \quad . \end{aligned}$$

Da questi studi si può vedere come l'applicazione di metodi multidimensionali alla selezione degli eventi rari sia da consigliare. È da tenere presente che la scelta di un metodo piuttosto che un altro dev'essere fatta di volta in volta a seconda del sistema studiato; inoltre l'applicazione non dev'essere effettuata ciecamente su tutte variabili a disposizione, ma si deve operare comunque una scelta di quelle più opportune.

L'applicazione dei metodi multivariati allo studio di altri decadimenti è possibile come ulteriore sviluppo di questo lavoro. La messa in opera degli esperimenti all'acceleratore LHC, prevista per il 2008, dunque, pone grande interesse sullo sviluppo di queste analisi, che potrebbero portare ai primi risultati già entro un anno dall'inizio della presa dati.

Bibliografia

- [1] Jogesh C. Pati and Abdus Salam. Unified lepton-hadron symmetry and a gauge theory of the basic interactions. *Phys. Rev.*, D8:1240–1251, 1973.
- [2] Jogesh C. Pati and Abdus Salam. Is baryon number conserved? *Phys. Rev. Lett.*, 31:661–664, 1973.
- [3] Jogesh C. Pati and Abdus Salam. Lepton number as the fourth color. *Phys. Rev.*, D10:275–289, 1974.
- [4] G. Valencia and S. Willenbrock. Quark-lepton unification and rare meson decays. *Phys. Rev.*, D50:6843–6848, 1994. hep-ph/9409201.
- [5] A. V. Kuznetsov and N. V. Mikheev. Vector leptoquarks could be rather light? *Phys. Lett.*, B329:295–299, 1994. hep-ph/9406347.
- [6] Sacha Davidson, David C. Bailey, and Bruce A. Campbell. Model independent constraints on leptoquarks from rare processes. *Z. Phys.*, C61:613–644, 1994. hep-ph/9309310.
- [7] (Belle Collaboration) Chang, M.C. Search for $B^0 \rightarrow l^+l^-$ at the Belle detector. *Phys. Rev. D*, 68(11):111101, 2003.
- [8] (CDF Collaboration) Abe, F. Search for the Decays $B_s^0, B_d^0 \rightarrow e^\pm \mu^\mp$ and Pati-Salam Leptoquarks. *Phys. Rev. Lett.*, 81(26):5742–5747, 1998.
- [9] P. Ball et al. B decays at the lhc. 2000. hep-ph/0003238.

- [10] D Martinez, JA Hernando, and F Teubert. LHCb potential to measure/exclude the branching ratio of the decay $B_s \rightarrow \mu^+\mu^-$, Apr 2007.
- [11] T. M. Aliev, A. Ozpineci, and M. Savci. $B_q \rightarrow l^+l^-$ gamma decays in light cone QCD. *Phys. Rev.*, D55:7059–7066, 1997. hep-ph/9611393.
- [12] T. M. Aliev, A. Ozpineci, and M. Savci. Rare $B \rightarrow l^+l^-\gamma$ decay and new physics effects. *Phys. Lett.*, B520:69–77, 2001. hep-ph/0105279.
- [13] Gad Eilam, Cai-Dian Lu, and Da-Xin Zhang. Radiative dileptonic decays of b mesons. *Phys. Lett.*, B391:461–464, 1997. hep-ph/9606444.
- [14] B. Aubert et al. Search for the Decays $B^0 \rightarrow e^+e^-\gamma$ and $B^0 \rightarrow \mu^+\mu^-\gamma$. 2007. arXiv:0706.2870 [hep-ex].
- [15] O. Schneider and T. Nakada. Lhcb trigger. Prepared for 4th International Conference on B Physics and CP Violation (BCP 4), Ago Town, Mie Prefecture, Japan, 19- 23 Feb 2001.
- [16] W.-M. Yao, C. Amsler, D. Asner, R.M. Barnett, and J. Beringer. Review of Particle Physics. *Journal of Physics G*, 33:1+, 2006.
- [17] *LHCb Technical Proposal*. CERN-LHCC-98-004.
- [18] *LHCb reoptimized detector design and performance Technical Design Report*. CERN-LHCC-2003-030.
- [19] *LHCb VELO (VERtEx LOcator) Technical Design Report*. CERN-LHCC-2001-011.
- [20] *LHCb RICH Technical Design Report*. CERN-LHCC-2000-037.
- [21] *LHCb inner tracker Technical Design Report*. CERN-LHCC-2002-029.
- [22] *LHCb magnet Technical Design Report*. CERN-LHCC-2000-007.
- [23] *LHCb outer tracker Technical Design Report*. CERN-LHCC-2001-024.

-
- [24] *LHCb calorimeters Technical Design Report*. CERN-LHCC-2000-036.
- [25] *LHCb muon system Technical Design Report*. CERN-LHCC-2001-010.
- [26] *Addendum to the Muon System Technical Design Report*. CERN-LHCC-2001-010.
- [27] *Second Addendum to the Muon System Technical Design Report*. CERN-LHCC-2005-0012.
- [28] *LHCb trigger system : Technical Design Report*. CERN-LHCC-2003-031.
- [29] *LHCb computing Technical Design Report*. CERN-LHCC-2005-019.
- [30] Per una descrizione del pacchetto EvtGen si consulti il sito:
<http://www.slac.stanford.edu/~lange/EvtGen/>.
- [31] Dean Karlen. Using projections and correlations to approximate probability distributions. *Computers in Physics*, 12:380, 1998. physics/9805018.
- [32] A. R. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7:179–188, 1936.
- [33] Giovanni Punzi. Sensitivity of searches for new signals and its optimization, 2003. physics/0308063.
- [34] F. Rademakers R. Brun. ROOT. Applicazione disponibile online presso <http://root.cern.ch/>.
- [35] A. Hocker, P. Speckmayer, J. Stelzer, F. Tegenfeldt, H. Voss, and K. Voss. TMVA. Software disponibile online presso <http://tmva.sourceforge.net/>.
- [36] A. Hocker et al. TMVA - Toolkit for Multivariate Data Analysis - User Guide, 2007. physics/0703039.

-
- [37] W. Bonivento and N. Serra. LHCb sensitivity to the LFV $B_{d,s}^0 \rightarrow e^\pm \mu^\mp$ decays. (LHCb-2007-028. CERN-LHCb-2007-028).
- [38] LHCb Bookkeeping Facility. Disponibile online presso <http://lhcbdata.home.cern.ch/lhcbdata/bkk/>.
- [39] Si veda il programma “bayes.f” alla pagina internet: <http://www-cdf.fnal.gov/physics/statistics>.
- [40] LHCb Collaboration. Rare Decays Twiki Page. <https://twiki.cern.ch/twiki/bin/view/LHCb/RareDecays>.

Ringraziamenti

Innanzitutto voglio ringraziare il Dott. Walter Bonivento, per avermi seguito in questo studio, per l'aiuto prezioso, le sempre interessanti discussioni e il supporto. Grazie Walter!

Ringrazio il Prof. Biagio Saitta per avermi dato la possibilità di lavorare sotto la sua supervisione, per le opportunità datemi e per i preziosi consigli.

Un grosso ringraziamento va a Nicola per l'aiuto in questo lavoro, e per i caffè e le risate.

Un GRAZIE particolare agli amici

Riccardo... anche per avermi ribadito la mia posizione su tante cose...,

Paolo... anche per non trattenere le risate nei momenti sbagliati...,

Nicola... anche per i discorsi improbabili...,

Andrea... anche per i semi di girasole ...,

ma soprattutto perché ci siete quando c'è bisogno.

Ringrazio Marta, Elena, Francesco, Antonio, Giamba e Andrea per l'amicizia e la compagnia.

Ringrazio i Damacreola per essere esistiti.

Ringrazio tutti gli amici e le amiche, che forse non leggeranno mai questa pagina... ma che mi hanno aiutato molto. Vi abbracerò di persona.

Ringrazio infine mio padre e mia madre, Giovanna, Mariella e Massimo e Fabio. Per tutto ciò che fanno per me, ma soprattutto per essermi vicini sempre. A loro sono dedicati gli sforzi di questa tesi.

