



Università degli Studi di Trieste

FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI
Corso di Laurea in Fisica

TESI DI LAUREA TRIENNALE

Esempi di misure di efficienza con il metodo Tag and Probe

Candidato:
Nicola Fontana

Relatore:
Dott. Giuseppe Della Ricca
Correlatore:
Dott. Andrea Schizzi

Alla Tre-Cani

Ringraziamenti

Colgo l'occasione offertami da questa tesi per ringraziare le persone che durante questi anni di studio mi hanno in qualche modo sostenuto e a volte sospinto.

Ringrazio mamma e papà per aver supportato la non facile scelta di avere un fisico in casa, ma soprattutto per ciò che fin da piccolo mi hanno regalato: una famiglia meravigliosa.

Ringrazio mio fratello per essere stato, ed essere ancora, il mio punto di riferimento, con la sua lungimiranza e il suo pragmatismo mi ha sempre fornito il supporto di cui spesso ho avuto bisogno.

Ringrazio tutti i parenti che hanno sempre avuto interesse per i miei studi e mi hanno a loro modo fatto sentire importante.

Ringrazio tutti i miei compagni di fisica, senza di loro questa esperienza non sarebbe stata così bella: Tina, Leti, Agnes, Prandalf, Enrico, Fisa, Trudi, Tommy, Pepi, Ghil, Fede, Matilde, Peppe, David, le coinquiline Amanda e Laura, Elena e anche i nuovi arrivati Fabio e Gabriele (spero di non aver tralasciato nessuno).

Ringrazio in particolar modo coloro che lungo questo cammino mi hanno davvero cambiato, Luca ed Oscar, dall'uno ho imparato la pazienza e l'organizzazione, dall'altro ho imparato che ci sono altri come me che vivono di sogni.

Ringrazio infine tutti gli amici del Comelico che mi hanno sempre fatto il tifo, in particolar modo Sergio che da quando ho coscienza di essere al mondo è sempre stato con me.

Indice

Introduzione	iv
1 LHC	1
1.1 Sistema di coordinate in uso ad LHC	3
1.2 Esperimenti ad LHC	4
1.2.1 CMS: <i>Compact Muon Solenoid</i>	5
1.2.2 ECAL: calorimetro elettromagnetico	9
2 Acquisizione e selezione dati a CMS	12
2.1 Sistemi di <i>Trigger</i>	12
2.1.1 <i>L1 trigger</i>	13
2.1.2 <i>High Level Trigger</i>	14
2.2 Selezione e ricostruzione di elettroni	14
3 Misure di efficienza con il metodo <i>Tag and Probe</i>	19
3.1 Fit dei dati	20
3.2 Studi di efficienza	22
3.2.1 Studi specifici di efficienza	23
Conclusioni	31

Introduzione

Lo studente realmente studioso è un solitario

H. Thoreau

In un esperimento di collisione quando si vuole selezionare un determinato evento si devono utilizzare dei criteri di selezione specifici, da applicare al campione di dati. Questi devono essere in grado di eliminare dal campione gli eventi di fondo non interessanti. Chiaramente questi strumenti possono essere opportunamente tarati in modo da garantirci diversi gradi di efficienza nella selezione. L'obiettivo di questa tesi è quello di validare il metodo *Tag and Probe*, un algoritmo in grado di misurare l'efficienza dei criteri di selezione. A tal fine abbiamo utilizzato il nostro algoritmo su un pacchetto di selezioni di cui già conoscevamo l'efficienza.

Questo insieme di tagli è stato applicato ai dati raccolti dall'esperimento CMS (*Compact Muon Solenoid*) presso l'acceleratore LHC (*Large Hadron Collider*) di Ginevra.

Le selezioni da noi analizzate sono l'ultimo passo di una lunga serie di raffinazioni dell'output proveniente da un esperimento di collisione. Vengono effettuate al fine di eliminare il rumore di fondo presente nei dati per ottenere un campione più pulito, da cui risultati più agevole acquisire informazioni sulla fisica del processo. Il metodo utilizzato per misurare questa efficienza, il *Tag and Probe*, permette di calcolare l'efficienza direttamente dal campione di dati. Questo tipo di approccio si presenta come alternativa al metodo Monte Carlo che invece si basa sulla ricostruzione virtuale dell'esperimento e la sua successiva simulazione. Il *Tag and Probe* verrà studiato sullo standard di selezioni WP80, messo appunto dai fisici di CMS al fine di ottenere un'efficienza di selezione dell'80%. I dati utilizzati per queste misure come detto provengono dall'esperimento CMS. Il campione è stato ottenuto nell'ambito della ricerca del bosone Z in eventi di collisione $p\text{-}p$ all'energia di 7 TeV, mediante il calcolo della massa invariante dei suoi prodotti di decadimento. Nel nostro lavoro la massa invariante del bosone Z è stata ricostruita mediante la ricerca di coppie elettrone-positrone e^+e^- . Questa scelta è stata fatta in quanto questo canale di decadimento è tra tutti il più semplice da osservare. Come primo passo è stata effettuata una misura globale dell'efficienza, basata su tutto il campione di dati. In seguito abbiamo deciso di analizzare nello specifico la dipendenza dell'efficienza da alcune variabili in particolare, per mettere alla prova la validità del metodo utilizzato. Al fine di ottenere un quadro più completo siamo andati ad analizzare anche un campione di dati generato con simulazione Monte Carlo. Questa analisi in parallelo ci ha permesso di avere sempre un riferimento da cui ottenere conferme sulla validità dei nostri risultati.

Capitolo 1

LHC

Il *Large Hadron Collider* (LHC) è la più grande ed innovativa macchina acceleratrice di particelle fino ad ora costruita, è progettata per accelerare protoni e ioni di piombo. La sua realizzazione ad opera della *European Organization for Nuclear Research* è iniziata nel 1998 e terminata nel 2008 con un costo complessivo di oltre due miliardi di euro.

L'acceleratore si trova nei pressi di Ginevra, al CERN, sul confine tra Francia e Svizzera. È stato realizzato sfruttando il tunnel sotterraneo precedentemente usato dall'esperimento LEP (*Large Electron-Positron Collider*), un anello sotterraneo di 27 Km ad una profondità media di 100 m. La macchina sarebbe dovuta entrare in funzione nel 2007, in realtà i primi protoni furono fatti circolare il 10 settembre 2008.

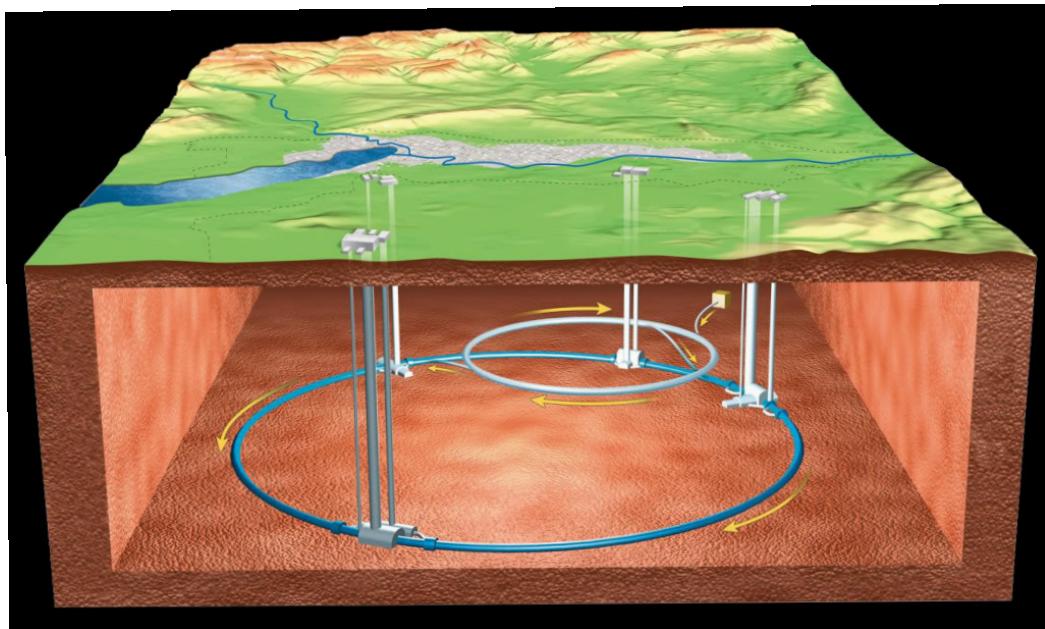


Figura 1.1: Anello sotterraneo di LHC

LHC è una macchina acceleratrice a fasci incrociati, ovvero una macchina in cui vengono accelerati contemporaneamente due fasci di particelle che circolano in direzioni opposte all'interno dell'anello. Una volta raggiunta l'energia d'interesse questi vengono fatti collidere in quattro punti. Per ottenere i migliori risultati possibili in termini energetici le particelle ad LHC viaggiano all'interno di tubi ad *ultra-alto-vuoto* i quali permettono di evitare che esse interagiscano con altri elementi presenti nell'aria. I fasci sono poi guidati lungo l'anello grazie a dei potenti magneti superconduttori, ovvero magneti costituiti da spire che operano in regime di superconduttevità. Tale regime di funzionamento richiede che le spire siano mantenute a

temperature di 1.9 K grazie ad un sistema di raffreddamento ad elio (3He) in condizioni di superfluidità. Tra le strutture che guidano le particelle nell'acceleratore vi sono 1232 dipoli magnetici che generano un campo di 8.36 T curvando il fascio e 392 quadrupoli magnetici che si occupano della sua focalizzazione. Nei pressi dei quattro punti di collisione sono presenti degli altri magneti che hanno il compito di strizzare il fascio in modo da massimizzare la probabilità di collisione.

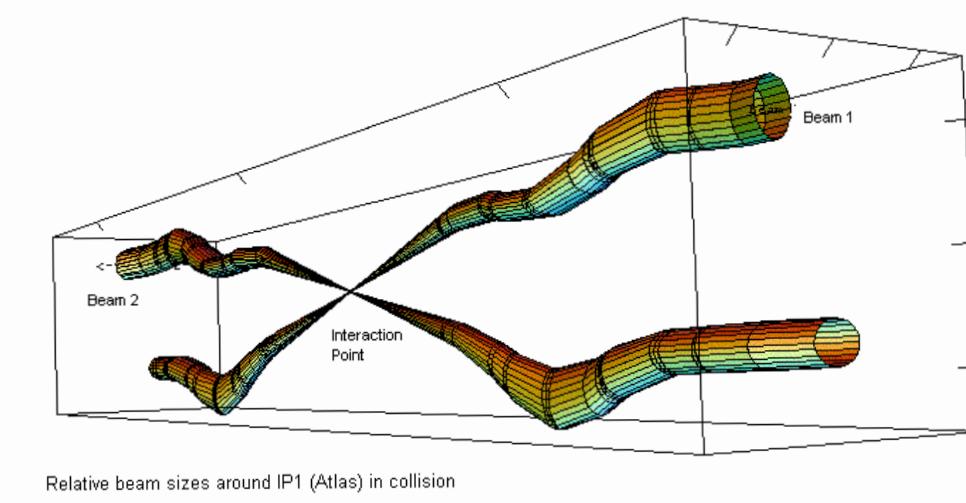


Figura 1.2: Viene mostrata la zona di collisione, è visibile la zona di pre-collisione in cui i fasci vengono "strizzati"

Prima di essere iniettati all'interno di LHC gli adroni o gli ioni di piombo devono essere accelerati ad energie dell'ordine dei GeV. Per ottenere tale energia al CERN è presente un complesso sistema di acceleratori con energie sempre crescenti utilizzati in sequenza. Ogni acceleratore aumenta l'energia delle particelle e le inietta direttamente nella macchina successiva e così sino all'immissione in LHC. Le particelle nei singoli acceleratori possono essere estratte per effettuare esperimenti ad energie minori, dal momento che quasi tutti possiedono una zona sperimentale adiacente. Vediamo ora la procedura necessaria per l'accelerazione dei protoni. Il primo passo viene compiuto dall'acceleratore lineare *Linac 2* che porta i protoni prodotti per ionizzazione dell'idrogeno ad un'energia di 50 MeV, questi vengono iniettati nel PS (*Proton Synchrotron Booster*) un acceleratore circolare che li accelera fino a 25 GeV. L'ultimo stadio prima dell'immissione in LHC viene compiuto da SPS (*Super Proton Synchrotron*) dove gli adroni raggiungono l'energia di 450 GeV. Questi poi vengono iniettati in LHC in entrambe le direzioni di circolazione, all'interno dei due tubi che viaggiano lungo l'anello, il processo richiede 4'20" per verso. L'ultimo stadio di accelerazione richiede 20" nel caso si debba raggiungere l'energia di 7 TeV. L'intero complesso di acceleratori è illustrato schematicamente in figura 1.3 a pagina seguente.

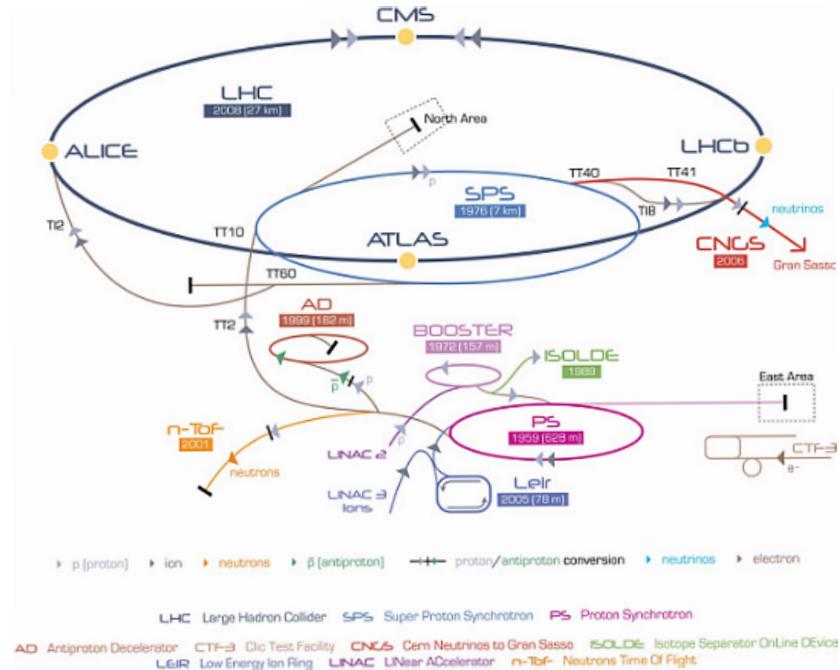


Figura 1.3: Rete di acceleratori del CERN

Una volta immessi in LHC i protoni viaggiano all'interno dell'acceleratore raggruppati in pacchetti detti *bunch* contenenti all'incirca 1.1×10^{11} elementi. Questi, vengono come anticipato strizzati prima del punto di collisione ad un diametro di $64 \mu\text{m}$. Una così alta densità di protoni per singolo *bunch* fa in modo che vi siano in media venti interazioni protone-protone per ogni *bunch-crossing*. La spaziatura temporale tra i *bunch* è di 25 ns e ciò fa sì che ben presto il numero di collisioni per secondo arrivi a a 600 milioni, $\sim 10^9 \text{ Hz}$. Una tale frequenza provoca un problema comune in questo tipo di esperimenti, il *pile-up*, ovvero viene generato un numero talmente alto di dati da renderne impossibile l'analisi ed il salvataggio senza l'utilizzo di opportuni filtri che in gergo si chiamano *trigger*.

L'energia massima raggiungibile ad LHC per fasci di protoni è di 14 TeV nel centro di massa, a queste energie gli adroni viaggiano a velocità molto prossime a quelle della luce (99,9999991%). Il *Large Hadron Collider* è in grado di accelerare anche ioni pesanti, ad esempio nuclei di piombo fino ad un'energia di 5.5 TeV nel centro di massa.

Con la costruzione di un acceleratore così avanzato si sono potute raggiungere energie fino a prima impensabili sulla Terra, e ciò ha aperto letteralmente le porte ad una nuova era per fisica nucleare, infatti energie così elevate consentono la verifica dell'esistenza o meno di alcune particelle ed alcuni fenomeni fisici che vengono predetti dal *Modello Standard* e che ancora oggi sono inspiegati.

1.1 Sistema di coordinate in uso ad LHC

Prima di parlare più in dettaglio degli aspetti strutturali del *Large Hadron Collider* e dei suoi esperimenti risulta necessario per la comprensione l'introduzione del particolare sistema di coordinate e dell'insieme di variabili cinematiche in uso presso questa struttura per descrivere gli eventi di collisione. Per ottenere la terna ortogonale destrorsa che viene utilizzata si immagini l'asse \hat{z} tangente alla linea di fascio nel punto di collisione, l'asse \hat{x} dato dall'intersezione del piano parallelo alla superficie terrestre e del piano perpendicolare al fascio (e dunque all'asse \hat{z}), con verso positivo diretto al centro dell'anello e l'asse \hat{y} perpendicolare ai due appena definiti.

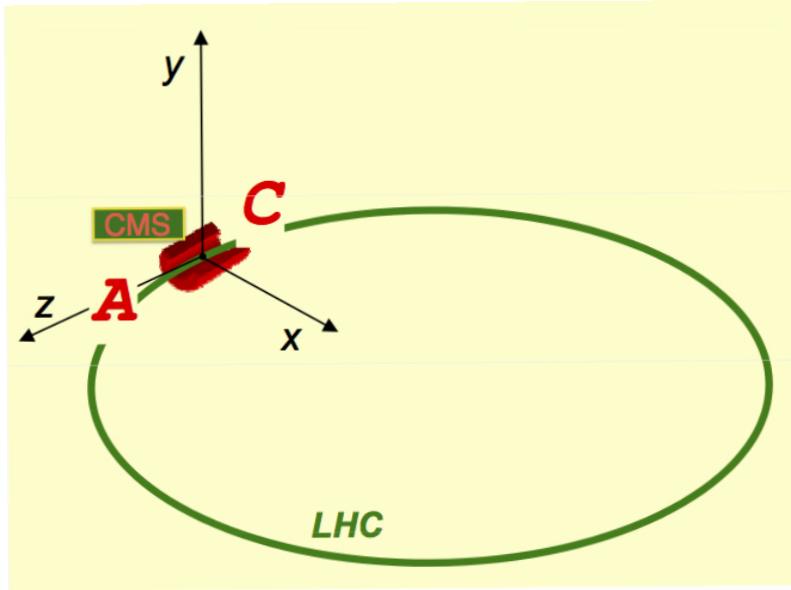


Figura 1.4: Terna ortogonale destrorsa in uso ad LHC

Ora introduciamo delle grandezze derivate dalla nostra terna:

$r = \sqrt{x^2 + y^2}$: rappresenta la distanza dalla linea di fascio.

$\theta = \arctan\left(\frac{y}{z}\right)$: rappresenta l'angolo polare.

$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right)$: viene detta *pseudorapidità*.

Dunque in realtà la vera terna di coordinate che viene utilizzata è (r, ϕ, η) , dove (ϕ, η) forniscono l'informazione circa la direzione della particella. Dal momento che in eventi di collisione adronici non è possibile risalire al momento iniziale per i partoni interagenti, si assume comunemente che la componente di momento perpendicolare alla linea di fascio sia trascurabile rispetto alla componente parallela e dunque si possa considerare nulla. Con queste ipotesi risulta utile la definizione di alcune variabili cinematiche rilevanti:

$E_T = \sqrt{p_T^2 c^2 + m^2 c^4} \simeq p_T c$: l'energia trasversa di una particella.

$p_T = p \sin\theta$ (GeV/c): il momento trasverso di una particella.

1.2 Esperimenti ad LHC

LHC come detto è un anello acceleratore, lungo tale anello vi sono quattro punti in cui i fasci vengono fatti collidere, attorno a queste zone sono stati costruiti i quattro esperimenti:

CMS : *Compact Muon Solenoid*.

ALICE : *A Large Ion Collider Experiment*.

ATLAS : *A Thoroidal Lhc Apparatus*.

LHCb : *Large Hadron Collider beauty*

vi sono inoltre altri due esperimenti siti in prossimità dei punti di collisione:

LHCf : *Large Hadron Collider forward*

TOTEM : *ToTal Elastic and diffractive cross section Measurement*

dediti all'analisi di particelle provenienti da interazioni a basso angolo di diffusione. Dei quattro esperimenti collocati sui punti di intersezione, CMS ed ATLAS sono quelli che presentano un programma di indagine fisica più ampio mentre ALICE e LHCb sono preposti a studi di carattere molto più specifico.

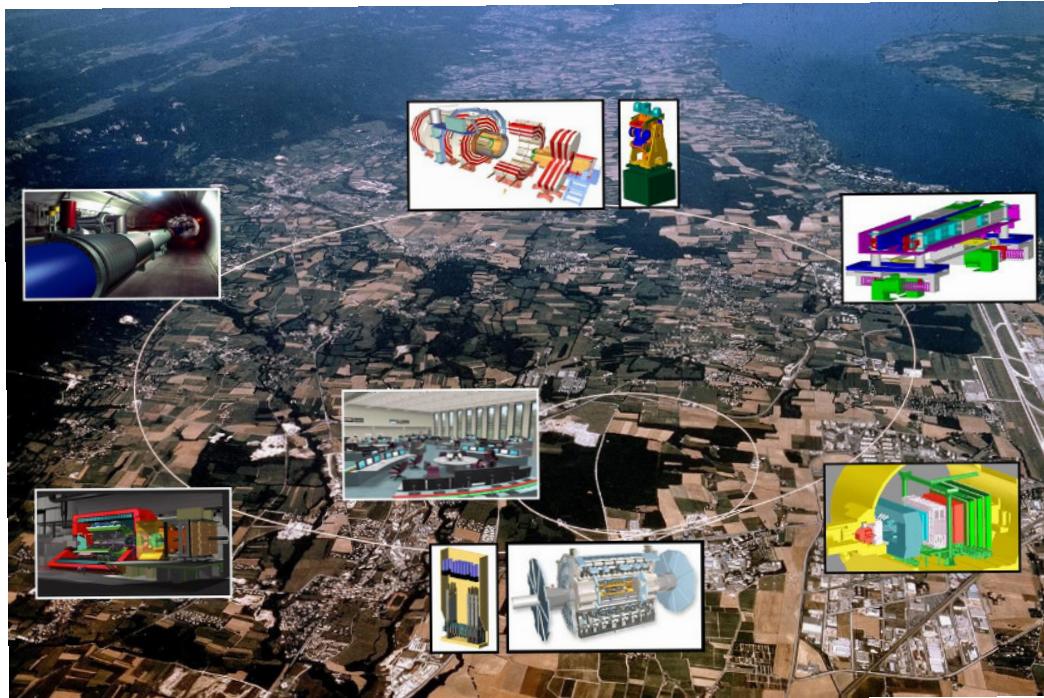


Figura 1.5: Vista aerea dell'intero complesso di LHC. Partendo dall'alto, in senso orario le immagini mostrano gli esperimenti: CMS, LHCf, LHCb, ATLAS e ALICE.

ALICE è progettato per l'analisi specifica di collisioni tra ioni pesanti, in particolare Pb-Pb ad energie di 2.76 TeV per nucleo, ci si attende che la temperatura e la densità di energia siano sufficientemente elevati da produrre un plasma di *quark-gluoni*. In questo nuovo stato della materia le componenti fondamentali della materia bariônica (quark e gluoni) sono in stato non legato. LHCb invece è ottimizzato per l'analisi e lo studio del quark *beauty*.

1.2.1 CMS: *Compact Muon Solenoid*

CMS è uno dei quattro esperimenti presenti sui punti di intersezione dei fasci, è situato sul suolo francese, precisamente a Cessy. Il rivelatore è collocato all'interno di una caverna sotterranea appositamente costruita e vanta dimensioni importanti: è lungo 21.6 m, ha un diametro di 14.6 m e pesa circa 12500 tonnellate. Le sue enormi dimensioni sono dovute al tipo di lavoro a cui è destinato, ossia deve essere in grado di rilevare le particelle prodotte nelle collisioni dei fasci di LHC, che possedendo energie molto elevate richiedono una grande quantità di materiale per poter essere fermate. In aggiunta a questo, avere un rivelatore di grandi dimensioni aumenta decisamente la precisione delle misurazioni in quanto è possibile

tracciare le traiettorie su distanze più estese. A differenza degli altri rivelatori non è stato costruito direttamente in loco, si è preferito procedere alla sua costruzione in superficie, mentre la caverna in cui sarebbe stato collocato era ancora in fase di realizzazione. Una volta ultimate le strutture che lo componevano queste sono state calate e assemblate all'interno della caverna. CMS è stato progettato in maniera modulare, dimodochè risulti più agevole la manutenzione o l'eventuale sostituzione di una sua componente. Come già anticipato CMS assieme ad ATLAS è l'esperimento ad LHC che presenta lo spettro di ricerca più ampio. È stato costruito, come ATLAS, allo scopo di indagare sulle *dimensioni extra*, su particelle che potrebbero costituire la *materia oscura*, per avere conferme sulla *supersimmetria* e sul *bosone di Higgs*, su cui ha già fornito dei risultati notevoli.

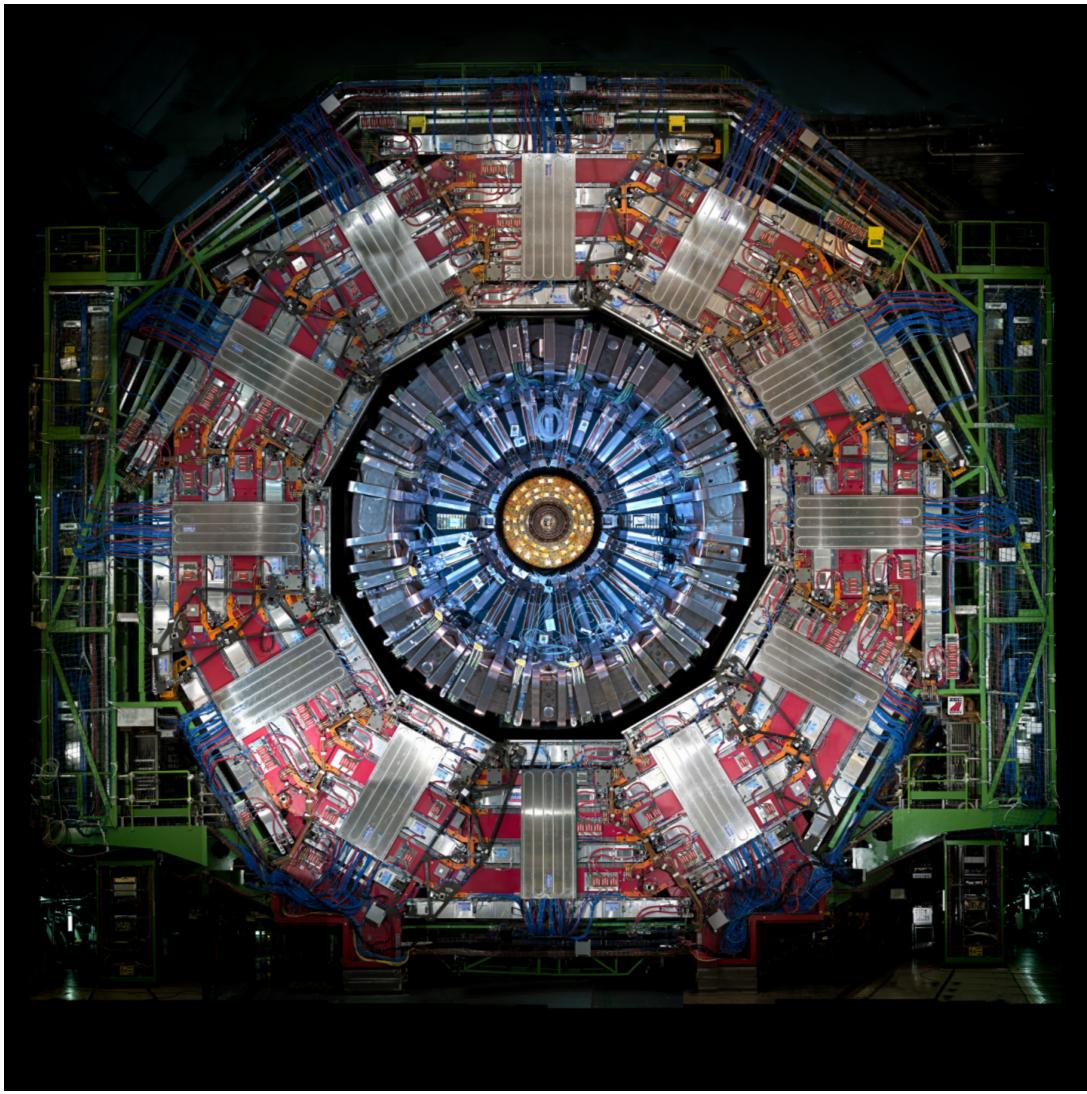


Figura 1.6: Vista frontale di CMS

La differenza tra ATLAS e CMS sta nella diversa struttura e i diversi rivelatori che possiedono, questa differenziazione è stata fatta volutamente allo scopo di avere la più ampia gamma di metodi di misura per lo stesso tipo di evento. La struttura di CMS è formata da due componenti principali, un corpo centrale cilindrico detto *barrel* e due tappi che chiudono il cilindro detti *endcaps*.

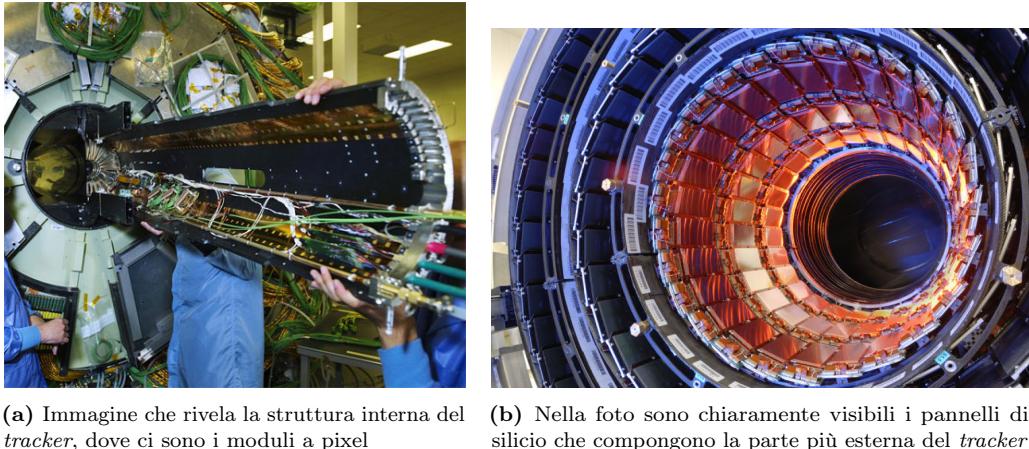


Figura 1.7

Il corpo centrale, il *barrel*, è costituito da quattro differenti rivelatori situati uno dentro l'altro, a partire dall'interno:

Tracker : è il tracciatore, posto in prossimità del punto di collisione in modo da ottenere la massima risoluzione sulla posizione dei vertici dell'interazione, ha il compito di rilevare la traiettoria delle particelle cariche interferendo con esse il meno possibile. Per soddisfare tale requisito esso è in grado di ricostruire le traiettorie utilizzando poche misurazioni con una precisione dell'ordine del μm . Dal momento che il tracciatore è l'elemento più prossimo al punto di interazione esso deve essere in grado di resistere ad alti dosaggi di radiazione senza compromettere la sua efficienza. Il *Tracker* costituito da due strati differenti di rivelatori:

- quelli più interni sono costituiti da tracciatori a pixel di silicio posizionati attorno alla linea di fascio in tre strati tubolari a distanze di 4, 7 e 11 cm chiusi da due dischi. Hanno le dimensioni di $100 \mu\text{m} \times 150 \mu\text{m}$, per ottenere la risoluzione spaziale massima ed evitare eventi di *pile-up*. Fondamentalmente il loro meccanismo di funzionamento è semplice: quando una particella colpisce il rivelatore un elettrone fuoriesce dall'atomo di silicio creando una coppia buca-elettrone che viene rilevata dall'elettronica di lettura. *Barrel* ed *endcap* sono ricoperti rispettivamente da 786 e 672 moduli a pixel
- quelli più esterni sono rivelatori a *microstrip* di silicio. La struttura del complesso a *microstrip* è formata da quattro strati interni al *barrel* (TIB) assemblati con due strati interni all'*endcap* (TID), la parte esterna del *barrel* è composta invece dai sei strati concentrici (due a singola faccia e due a doppia faccia), per chiudere il tutto ci sono i due *endcap* (TEC). In totale (*barrel-endcap*) sono presenti 15400 moduli con un totale di 10 milioni di *microstrip* collegati a 80000 chip di lettura. Il complesso di questi rivelatori riesce a coprire la regione di spazio $\eta \leq 2.4$

ECAL : è il calorimetro elettromagnetico, situato attorno al *tracker*, si occupa di fermare le particelle scaturite dall'interazione con lo scopo di misurare l'energia da esse depositata e riconoscerne dunque la *segnatura*. ECAL per come è costruito è in grado di rilevare fotoni elettroni e positroni

HCAL : è il calorimetro adronico, circonda il calorimetro ECAL, e come esso si occupa di fermare le particelle prodotte nella collisione. Le particelle che rileva HCAL sono quelle che sfuggono ad ECAL ovvero gli adroni, ad esempio: neutroni, protoni, kaoni e pioni

Rivelatori per i muoni : sono collocati all'esterno del magnete superconduttore alternati da strutture che concentrano il campo magnetico di ritorno del magnete solenoidale. La scelta di collocare questi rivelatori così lontano dal centro dell'interazione è figlia del fatto che i muoni interagiscono molto debolmente con la materia e dunque possono attraversarne grossi strati senza essere alterati. Questi rivelatori sono di tre differenti tipologie:

- *Drift Tubes* : tubi a deriva atti alla misurazione accurata delle traiettorie nella zone del *barrel*
- *Resistive Plate Chambers* : sono caratterizzate da un tempo di risposta estremamente rapido e sono presenti sia nel *barrel* che nell'*endcap*
- *Cathode Strip Chambers*: sono presenti esclusivamente nell'*endcap*, anche essi con il compito di tracciare con precisione le traiettorie.

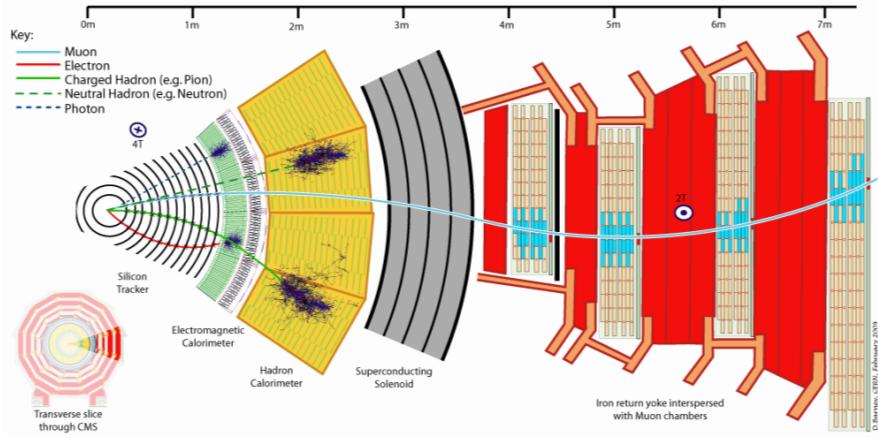


Figura 1.8: Immagine che mette in evidenza la sequenza di rivelatori dall'interno verso l'esterno di CMS.

I primi tre rivelatori (*tracker*, ECAL ed HCAL) si trovano racchiusi all'interno della componente attorno a cui è stato costruito tutto il progetto, il magnete solenoidale superconduttore, mentre le camere a muoni sono all'esterno. Il magnete superconduttore è un cilindro avvolto da spire superconduttrici mantenute a temperatura di 4.6 K e pesa 12000 tonnellate.

È il più grande magnete superconduttore fin ora costruito ed ha la funzione fondamentale di produrre il campo magnetico necessario (3.8 T ovvero 100000 volte quello generato dalla Terra) al sistema di tracciamento per la misura del momento e del segno delle particelle che possiedono carica. La necessità di un magnete così potente deriva dal bisogno di rilevare particelle con un momento molto alto come ad esempio i muoni, questo perché più una particella carica è veloce e meno viene curvata dal campo magnetico, ed è proprio dall'entità della curvatura che si determina il momento. La presenza di tutti i diversi rivelatori è indispensabile per poter distinguere tutte le particelle che vengono prodotte in un evento di collisione ad esempio :

- i fotoni sfuggono al tracciatore e vengono rilevati solo grazie ad un segnale che rilasciano all'interno del calorimetro elettromagnetico
- positroni ed elettroni lasciano anche loro un deposito nel calorimetro elettromagnetico ma vengono pure rilevati dal tracciatore che ne rileva le traiettorie curve

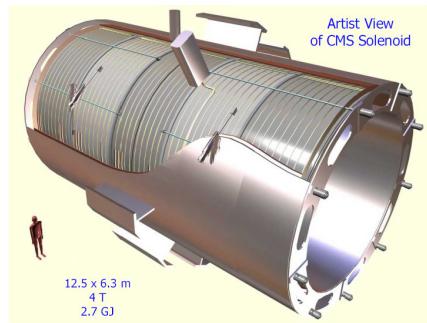


Figura 1.9: Il magnete solenoidale di CMS

- adroni carichi che vengono rilevati dal tracciatore e in seguito depositano tutta la loro energia all'interno del calorimetro adronico
- adroni neutri che ovviamente non vengono tracciati, non essendo carichi, ma lasciano un deposito energetico nel calorimetro adronico.

Dunque, riassumendo, CMS è un rivelatore di ultima generazione, costruito per tenere traccia di tutti i segnali prodotti in eventi di collisione grazie ad una serie di rivelatori sofisticati che svolgono compiti complementari per l'identificazione delle particelle prodotte.

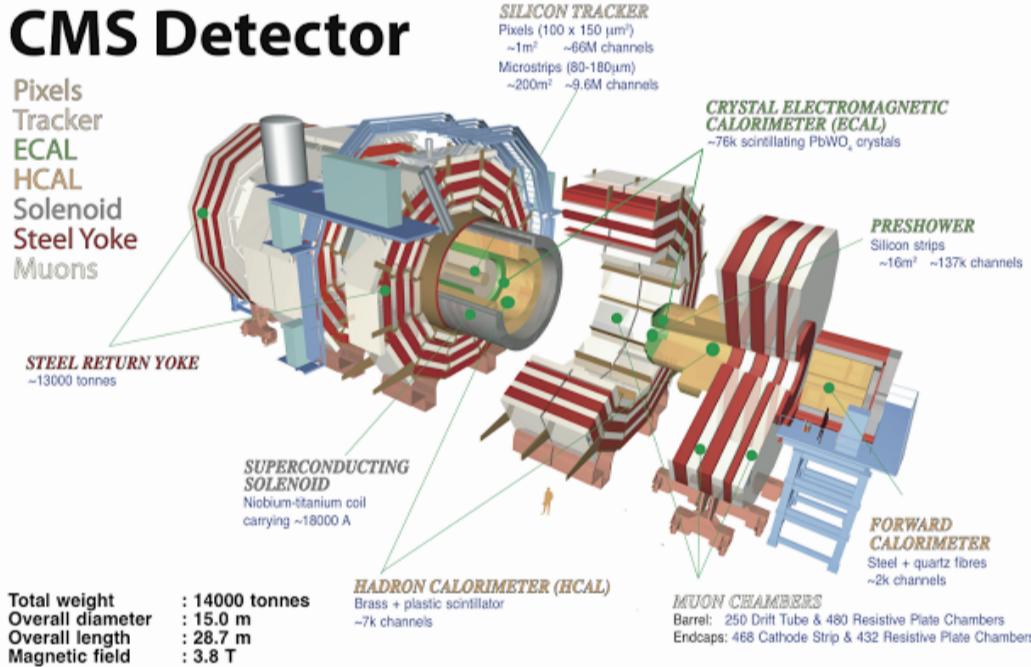


Figura 1.10: Sezione tridimensionale che illustra schematicamente la struttura di CMS.

1.2.2 ECAL: calorimetro elettromagnetico

Andiamo ora a vedere più da vicino le peculiarità del calorimetro elettromagnetico ECAL [1] dal momento che le misure da noi effettuate si basano su coppie *elettrone-positrone* individuate proprio grazie a questo rivelatore. Come già citato il compito fondamentale per un calorimetro è quello di frenare le particelle che lo attraversano fino a fermarle, al fine di rivelare dei segnali che le identifichino. Per ottenere ciò un calorimetro deve essere composto da un materiale assorbitore con cui le particelle prodotte nella collisione possano facilmente interagire. Quando particelle ad altra energia entrano nel calorimetro interagiscono con l'assorbitore dando vita a sciami elettromagnetici costituiti da particelle secondarie le quali vengono poi assorbite dal calorimetro. Il punto di forza di questo principio di funzionamento sta nel fatto che maggiore è l'energia della particella, maggiore è la precisione con cui essa viene misurata.

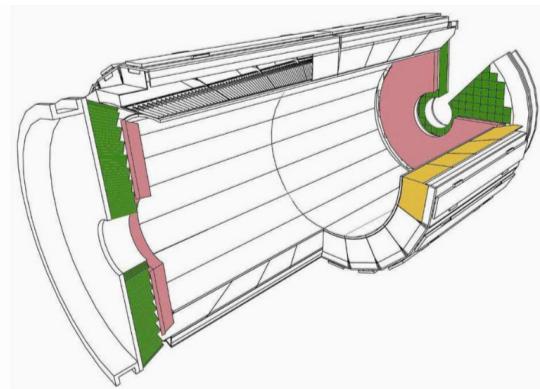


Figura 1.11: Visione 3-d di ECAL, sono messi in evidenza sugli *endcaps* i rivelatori *preshower* in rosa, i supermoduli dell'*endcap* in verde e quelli del *barrel* in giallo.

Dal momento che ECAL è votato alla cattura di elettroni, positroni e fotoni, particelle che interagiscono in modo elettromagnetico con la materia, è necessario che il materiale assorbitore sia composto da elementi ad alto numero atomico Z. Chiaramente fermare la particella che deve essere misurata non è sufficiente occorre che essa venga rilevata. Il modo in cui il segnale generato dall'interazione della particella con il rivelatore viene prodotto distingue due categorie differenti di calorimetri:

a campionamento : in questa tipologia di calorimetro vengono alternati a strati di materiale assorbitore degli strati di materiale, detto attivo, in grado di reagire al passaggio della particella con la produzione di un segnale

omogenei : in questa classe di calorimetri invece lo stesso materiale ha la funzione di assorbitore e di mezzo attivo.

La seconda categoria è quella in cui rientra ECAL, il materiale atto a svolgere la duplice funzione di assorbitore-scintillatore è il *tungstenato di piombo* PbWO₄, il quale ha la caratteristica di emettere la maggior parte della sua luce di scintillazione ($\sim 80\%$) in un intervallo temporale di 25 ns, che è proprio il tempo che segna la spaziatura dei *bunch* di LHC.

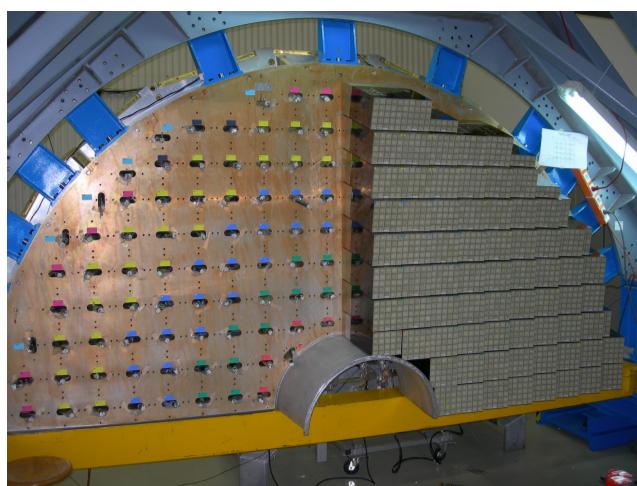


Figura 1.12: Porzione di *endcap* in cui sono visibili i cristalli di tungstenato organizzati in supermoduli.

Questo materiale inoltre presenta un elevata densità (8.28 g/cm^3) necessaria per avere una ridotta lunghezza di radiazione ($\chi_0 = 0.89 \text{ cm}$). Questa soluzione permette di avere dimensioni contenute del calorimetro e ciò è senza dubbio un grosso vantaggio. Il prezzo da pagare però è la debolezza del segnale emesso da questo materiale speciale. Per rendere riconoscibile il segnale prodotto dal tungstenato è necessario l'utilizzo di fotomoltiplicatori creati appositamente allo scopo. Il calorimetro è composto da un alto numero di cristalli al fine di ottenere la granularità necessaria alle misure che dovrà effettuare, tali cristalli soddisfano inoltre al requisito di alta resistenza alle massicce dosi di radiazione causate dalle collisioni alle energie previste per LHC. La geometria del calorimetro ricalca quella dell'intera struttura e così anche esso è formato da una parte cilindrica detta *barrel* e da due tappi detti *endcaps*. Il *barrel* è formato da 61200 cristalli, ognuno del peso di 1.5 Kg, a loro volta organizzati in trentasei cosiddetti *supermoduli*, mentre l'*endcap* ne presenta circa 15000. La regione definita dal *barrel* copre una ampiezza $\eta \leq 1.479$ mentre quella dell'*endcap* copre un ampiezza $1.479 < \eta < 3.0$. Nella parte interna degli *endcap* è collocato un rivelatore di *preshower* che serve a massimizzare l'identificazione di pioni neutri da parte del calorimetro e distinguere fotoni singoli di alta energia da coppie di fotoni a bassa energia. Per tradurre i segnali prodotti dai cristalli e amplificati dai fotomoltiplicatori è necessaria una complessa elettronica di lettura che non solo deve tradurre i segnali in dati, ma deve farlo in tempi brevissimi dato l'alto tasso di produzione di segnali dovuti alle collisioni. Deve inoltre operare in presenza di un forte campo magnetico generato dal magnete superconduttoro. Per svolgere al meglio questo compito l'elettronica di lettura è composta da fotodiodi a valanga per il *barrel* e per fotodiodi a vuoto per l'*endcap*, la diversa scelta per *barrel* ed *endcap* è dovuta al diverso flusso e tipo di particelle che attraversano le due componenti.

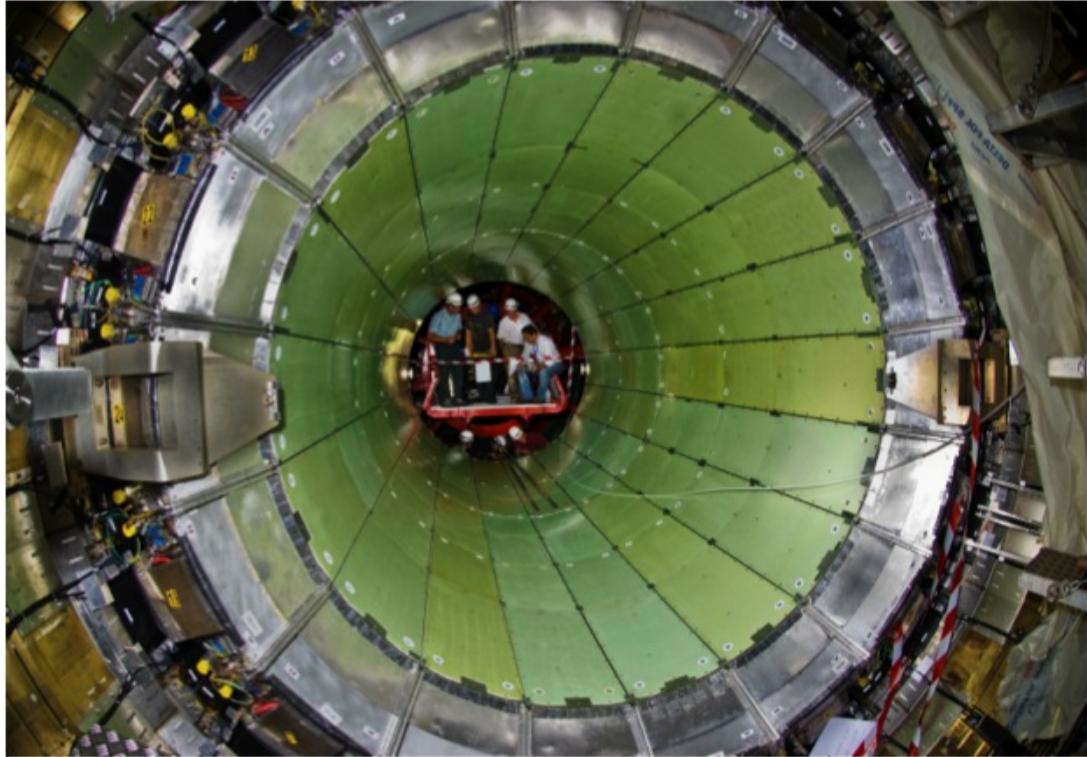


Figura 1.13: Il calorimetro ECAL collocato all'interno di CMS.

Capitolo 2

Acquisizione e selezione dati a CMS

Questo capitolo della tesi è dedicato ad uno degli aspetti più delicati in un esperimento di collisione, ovvero l'acquisizione e la selezione dei dati che verranno salvati e in seguito analizzati. Come abbiamo già anticipato le collisioni generate da LHC producono in media venti eventi per *bunch-crossing* con una frequenza di 25 ns. Questo flusso di dati corrisponde a circa 10^9 Hz, una tale frequenza produce un output troppo grande per poter essere registrato. Tuttavia non tutti i segnali prodotti sono relativi ad eventi rilevanti dal punto di vista fisico, possono esserci numerosi eventi relativi a urti non frontali a bassa energia che non siamo affatto interessati a salvare. Questa frequenza di dati va abbassata di un fattore 10^7 per ottenere un output attorno ai 100 Hz. I dati poi verranno ulteriormente elaborati da un centro di calcolo prima di essere salvati ed essere pronti per essere analizzati.

2.1 Sistemi di *Trigger*

Parliamo ora più concretamente degli strumenti veri e propri che operano le scelte necessarie per ridurre il flusso di informazioni prodotte negli eventi di collisione, i cosiddetti *trigger*. Il loro compito è quello di scartare segnali non interessanti al fine di ottenere un campione di dati ritenuto fisicamente rilevante. Chiaramente progettare un sistema in grado di compiere questa operazione, tenendo conto della frequenza di urti di LHC, dove addirittura nuovi eventi si sovrappongono a quelli precedenti, è una sfida non banale. Fino all'inizio degli anni '90 il sistema più avanzato in questo ambito era un *trigger* a tre livelli inaugurato la prima volta nell'esperimento *Zeus* nel 1985. Ogni stadio provvedeva alla riduzione della flusso di dati in modo che esso fosse adatto allo stadio successivo. Il processo filtraggio avveniva nel seguente ordine [2]:

1. i dati venivano filtrati dal primo stadio, L1, con un sistema hardware che operava le selezioni più generali per abbassare la frequenza a 30 KHz
2. l'output proveniente da L1 veniva raccolto dal secondo stadio, L2, dove veniva effettuata una parziale ricostruzione dell'evento su cui basare poi la selezione ulteriore da applicare, l'output in uscita arrivava ad una frequenza di 350 Hz
3. al terzo stadio, L3, una *computer-farm* si occupava di effettuare le ultime selezioni per ottenere un output a 100 Hz pronto per essere salvato per successive analisi.

Per l'esperimento CMS però i responsabili del settore riguardante i *trigger* scelsero di discostarsi dal classico approccio a tre livelli e optarono per un sistema a due livelli, in sostanza veniva saltato lo stadio L2, dimodochè i dati passassero direttamente da L1 alla *computer-farm*. La scelta fu fatta con il chiaro intento di migliorare la qualità e la tempistica di questi *filtr*, L2 veniva considerato troppo limitante a causa della sua struttura hardware.



Figura 2.1: Veduta della farm *Tier-1* di LHC

Questo stadio d'altra parte era al tempo necessario in quanto non esisteva una banda in grado di reggere la mole di dati uscente direttamente da L1 per portarla alla *computer-farm*. Nonostante il problema di trovare la banda adatta a supportare il flusso di dati di L1 si decise comunque di procedere verso lo snellimento del sistema di *trigger*, così con l'elettronica al tempo disponibile, siamo nel 1995, venne proposto un sistema a due livelli in cui il solo 10% dei dati veniva spedito ai computer che avrebbero poi deciso se l'evento fosse interessante o meno. Fortunatamente nel 2000 l'elettronica necessaria venne sviluppata e fu possibile dunque l'implementazione completa di un sistema a due livelli. Questo sistema richiese un miglioramento dell'hardware di L1 per renderlo capace di produrre un output immediatamente fruibile ai computer della *farm*. Dunque il sistema di acquisizione dati presente a CMS è composto da due macroelementi:

Trigger L1 : è il primo livello, totalmente hardware, in grado di compiere decisioni in tempi molto ristretti riguardo all'importanza fisica degli eventi che analizza

High Level Trigger : è il secondo stadio, software, in grado di compiere ricostruzioni parziali e perciò effettuare decisioni più raffinate, ha il vantaggio considerevole di essere molto flessibile in quanto riprogrammabile e quindi adattabile a diversi esperimenti ma paga il tutto con un tempo di risposta notevolmente più dilatato rispetto al primo stadio.

2.1.1 *L1 trigger*

Il trigger L1 di CMS è progettato per il riconoscimento di elettroni, positroni, fotoni, muoni, getti adronici ed energia trasversa mancante. Deve essere altamente efficiente nonostante valori di soglia per l'energia molto bassi in modo da fornire una grande quantità di dati per l'analisi, inoltre deve essere altamente inclusivo dal momento che LHC ha spinto molto oltre i range di energia e ci si attende di osservare fenomeni nuovi ed inaspettati. Come abbiamo già anticipato è un *trigger* hardware, ovvero è costituito da circuiti logici appositamente studiati, in grado di rispondere in tempi molto brevi e per questo in grado di compiere solo decisioni piuttosto *semplici*. La spaziatura temporale tra i *bunch* di LHC è di 25 ns il che impone alcuni limiti piuttosto restrittivi sull'architettura di L1. Il tempo impiegato da quando avviene la

generazione del segnale nei rivelatori a quando il *trigger* prende la sua decisione è di $3.2 \mu s$ [3]. Ogni processo all'interno del *trigger* deve ripetersi ogni 25 ns e questo implica che ogni passo della sua analisi deve durare meno di tale lasso di tempo. Inoltre, per ottimizzare i tempi, informazioni specifiche vengono veicolate in punti specifici del sistema, dove verranno elaborate. La richiesta di una risposta così rapida, $3.2 \mu s$, fa sì che al *trigger* siano accessibili per l'analisi solo i dati riguardanti l'attimo immediatamente successivo alla collisione, che ovviamente sono solo una parte, ad esempio i dati provenienti dalle *Muon Drift Chambers* del *barrel* impiegano 400 ns prima di raggiungere L1. Per la stessa ragione non sono accessibili nemmeno i dati del *tracker* e quelli del rivelatore di *preshower*. Il *trigger* di primo livello non può pertanto sfruttare appieno la risoluzione dei rivelatori da cui riceve informazioni, ma deve limitarsi a lavorare con dati meno raffinati, che gli vengono inviati dai singoli sistemi di selezione presente sui rivelatori stessi, nel caso di ECAL questi sistemi sono rappresentati dalle *calotower* o torri calorimetriche. In conclusione L1 riduce la frequenza di dati di un fattore 10^3 .

2.1.2 High Level Trigger

L'*High Level Trigger* si occupa della seconda e ultima fase di selezione dati per quanto riguarda la parte *online*, è un filtro di tipo software che sfrutta ricostruzioni parziali degli eventi per prendere le sue decisioni. HLT elabora tutti gli eventi che hanno passato la selezione effettuata dal *trigger* di primo livello L1. Per ottimizzare l'efficienza, ovvero ridurre al minimo l'utilizzo di CPU, HLT deve essere in grado di eseguire ricostruzioni parziali con la minima quantità di dati necessaria ad applicare un pacchetto di selezioni che eliminino il fondo nei dati. Le ricostruzioni parziali sono effettuate sulla base delle informazioni fornite dal *trigger* di primo livello, ad esempio la ricostruzione di un elettrone parte con la ricerca di una torre calorimetrica, selezionata da L1 affinché abbia un deposito superiore a 5 GeV. Una volta individuata la torre calorimetrica indicata da L1, usando tutta la risoluzione di ECAL [4] viene prima individuato il cristallo con il contenuto energetico più consistente e poi, con un algoritmo detto ad *isola* [5], vengono individuati cristalli adiacenti con un contenuto energetico al di sopra di una determinata soglia. Questa procedure permette di individuare un *cluster* da associare ad un segnale nel rivelatore a pixel del *tracker*. Introducendolo abbiamo sottolineato come HLT sia andato a sostituire i *trigger* di secondo e terzo livello unendoli in uno stadio unico. Nonostante ciò, al suo interno esso agisce comunque in due fasi distinte. Al primo stadio HLT usa informazioni parziali provenienti dai calorimetri e dai rivelatori di muoni al secondo vengono utilizzate anche informazioni provenienti dal *tracker*. Alcuni suggeriscono di introdurre idealmente uno stadio intermedio, in tale fase vengono sfruttate principalmente le informazioni dei calorimetri ma per conferma è possibile che il *trigger* acceda anche a informazioni del tracciatore, appositamente selezionate dal *trigger* di primo livello. Va sottolineato che circa un 30% [6] del flusso di dati è riservato alla calibrazione dei trigger, così da assicurare un accurato monitoraggio delle loro prestazioni.

2.2 Selezione e ricostruzione di elettroni

Nell'analisi dati della presente tesi sono state analizzate coppie di elettroni e positroni provenienti dal decadimento :

$$Z^0 \rightarrow e^+ e^- \quad (2.1)$$

perciò è opportuno dare una breve descrizione del processo che ha portato al campione di dati utilizzato nel nostro lavoro. Si vanno ad analizzare i dati che hanno superato i due *trigger* L1 ed HLT e che sono stati salvati, per effettuare una ricostruzione completa dell'evento. Nel caso specifico degli elettroni ciò consiste nella *reclustering* dei cristalli del calorimetro elettromagnetico e nella completa ricostruzione delle tracce. Il passo successivo consiste nel raffinare ulteriormente i dati acquisiti, dove per raffinare si intende eliminare il rumore di fondo

al fine di ottenere un segnale più pulito. Per fare ciò sono stati messi appunto degli appositi pacchetti di tagli da applicare ai dati per ottenere una determinata efficienza nella selezione di elettroni. I diversi pacchetti sono stati etichettati sulla base delle efficienze di selezione che garantiscono, dette anche punto di lavoro o *working point*. Per il campione di dati su cui abbiamo lavorato, ottenuto dai *run* effettuati nel 2010, sono state messe appunto sei configurazioni [7] di tagli con diverse efficienze:

working point 95 : WP95

working point 90 : WP90

working point 85 : WP85

working point 80 : WP80

working point 70 : WP70

working point 60 : WP60

dove la cifra indica in percentuale la relativa efficienza (es. WP95 = efficienza del 95%). Tuttavia gli standard di selezioni che vengono più comunemente usati sono il WP95 ed il WP80. Nell'applicazione dei tagli quello che si cerca di fare come già detto è ridurre il rumore di fondo nei dati. Nel nostro caso (elettroni) questo corrisponde ad eliminare segnali provenienti da particelle che hanno una segnatura simile a quella dell'elettrone nel rivelatore ma anche a scartare elettroni provenienti da processi secondari (non prodotti direttamente nelle collisioni protone-protone). Questi pacchetti sono indispensabili poiché i segnali di fondo possono passare le selezioni dei *trigger* ed è quindi necessario che questi criteri di raffinazione dei dati lavorino per ottimizzare il rapporto tra segnale e rumore e rendere notevolmente più agevole l'analisi fisica. Proponiamo di seguito la tabella redatta a CMS nel 2010 per i sei standard elencati.

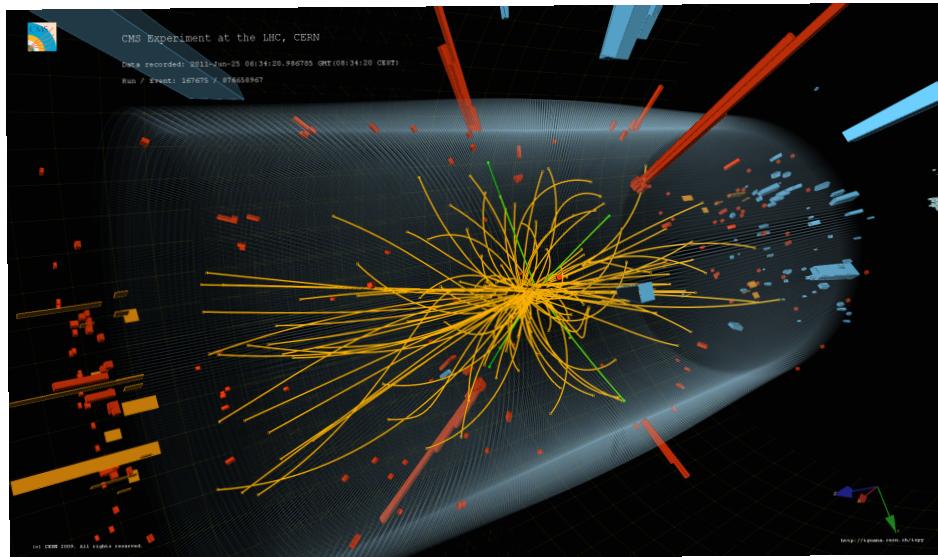
Efficiencies	95%	90%	85%	80%	70%	60%
Coverage Rejection						
missing hits \leq	1	1	1	0	0	0
Dist	N/A	0.02	0.02	0.02	0.02	0.02
$\Delta\cot\theta$	N/A	0.02	0.02	0.02	0.02	0.02
BARREL						
Isolation						
Combined Isolation						
combinedISO	0.15	0.10	0.09	0.07	0.04	0.03
Relative Isolation						
trackRel03	0.15	0.12	0.09	0.09	0.05	0.04
ecalRel03	2.00	0.09	0.08	0.07	0.06	0.04
hcalRel03	0.12	0.10	0.10	0.10	0.03	0.03
Electron ID						
σ_{min}	0.01	0.01	0.01	0.01	0.01	0.01
$\Delta\phi$	0.8	0.8	0.06	0.06	0.03	0.025
$\Delta\eta$	0.007	0.007	0.006	0.004	0.004	0.004
HoE	0.15	0.12	0.04	0.04	0.025	0.025
ENDCAPS						
Isolation						
Combined Isolation						
	0.1	0.07	0.06	0.06	0.03	0.02
Relative Isolation						

(Continua alla pagina successiva)

tracRel03	0.08	0.05	0.05	0.04	0.025	0.025
ecalrel03	0.06	0.06	0.05	0.05	0.025	0.02
hcalRel03	0.05	0.03	0.025	0.025	0.02	0.02
Electron ID						
σ_{min}	0.03	0.03	0.03	0.03	0.03	0.03
$\Delta\phi$	0.7	0.7	0.4	0.3	0.2	0.2
$\Delta\eta$	0.01	0.009	0.007	0.007	0.005	0.005
HoE	0.07	0.05	0.025	0.025	0.025	0.025

Tabella 2.1: Elenco dei tagli relativi ai sei standard di selezione

Andiamo ora a descrivere brevemente il significato di questi tagli. Come prima cosa notiamo che la tabella presenta due sezioni differenti, una per il *barrel* ed una per gli *endcap*. Una tale divisione è naturalmente dovuta al fatto che le due sezioni presentano caratteristiche differenti che causano prestazioni differenti e per tale motivo necessitano di tagli diversi. Ci sono inoltre tre principali categorie di variabili: *Conversion Rejection*, *Combined* e *Relative Isolation* ed *Electron ID*, andiamo ad analizzarle una per volta.

**Figura 2.2:** Evento di collisione ricostruito a CMS

Conversion Rejection Questa categoria di tagli è dedita alla verifica che gli elettroni rivelati non discendano dalla conversione di due fotoni che vanno a creare una coppia e^+e^- , un fenomeno che con i rate di produzione di particelle di LHC è tutt'altro che trascurabile. Fortunatamente le tracce di coppie e^+e^- generate da conversione di fotoni hanno una geometria molto particolare che permette di identificarle con relativa semplicità. Le tracce della coppia devono condividere un vertice ed appartenere al piano $r-\phi$ [7], questo porta a definire due quantità:

Dcot : è la differenza tra la cotangente di θ , l'angolo polare della traccia dell'elettrone selezionato e l'analogia quantità relativa all'elettrone che si suppone completare la coppia generata da conversione

Dist : è la distanza tra le tracce dei due elettroni quando queste diventano parallele.

Dal momento che la conversione di fotoni può avvenire solo dopo l'interazione con la materia, i segnali di elettroni prodotti da tale fenomeno avranno delle tracce che rispetto a quelle di elettroni prodotti nell'evento iniziano in strati più esterni del tracciatore.

Isolation In questa categoria rientrano i tagli utili a distinguere elettroni prodotti nell'interazione vera e propria da elettroni che hanno origine dai prodotti di decadimento di un processo di adronizzazione di un *gluone* o un *quark*. Ciò che rende distinguibile questo tipo di elettroni è l'alto numero di tracce nel *tracker* che li accompagnano, relative a particelle cariche copiosamente prodotte in processi di adronizzazione. In tal caso dunque si dice che l'elettrone non è *isolato*. Per quantificare l'isolamento si definiscono quattro quantità:

trackIso : rappresenta l'isolamento nel *tracker*, è la sommatoria su tutte le tracce che è stato possibile ricostruire divisa per il momento associato alla traccia dell'elettrone

$$\text{trackIso} = \frac{\sum p_T(i)}{p_T} \quad (2.2)$$

ecalIso : rappresenta l'isolamento all'interno di ECAL ed è la somma di tutti i depositi energetici compresi in un cono, che circonda il deposito energetico attribuito al possibile elettrone diviso per il deposito dell'elettrone. Il cono è definito da un'ampiezza $\Delta R = 0.3$ dove $R = \sqrt{\eta^2 + \phi^2}$

$$\text{ecalIso} = \frac{\sum E_T(i)}{E_T} \quad (2.3)$$

hcalIso : è l'isolamento all'interno di HCAL ed è anch'esso la somma dei depositi energetici all'interno del cono che circonda il deposito relativo al supposto elettrone diviso per il deposito dell'elettrone

$$\text{hcalIso} = \frac{\sum E_T(i)}{E_T} \quad (2.4)$$

CombinendIso : è semplicemente definito come

$$\text{combinedIso} = \frac{\sum_{\text{tracker}} p_T(i) + \sum_{\text{ECAL}} E_T(i) + \sum_{\text{hcal}} E_T(i)}{E_T} \quad (2.5)$$

Electron ID Questo pacchetto di selezioni serve a identificare segnali prodotti da elettroni nell'evento. In queste selezioni si vanno ad analizzare l'energia depositata dalla particella all'interno del *supercluster* di cristalli, l'allineamento di quest'ultimo con la traccia che è stata ricostruita nel *tracker* e la frazione di energia adronica nel *supercluster*. Queste osservazioni inducono l'introduzione di quattro variabili:

- σ_{min} : lungo η la traccia di un elettrone nel calorimetro presenta dispersione minima, perciò il deposito energetico lungo tale direzione deve essere concentrato in pochi cristalli, è dunque possibile impostare una soglia alla deviazione standard relativa alla distribuzione del deposito per eliminare tutte le particelle che presentano ampia dispersione in η all'interno del calorimetro.

- Δ_η, Δ_ϕ : queste due variabili riguardano la distanza tra la traccia che viene ricostruita dal *tracker* e il baricentro del *cluster* relativo al deposito energetico della particella, chiaramente per identificare un elettrone si richiede che il valore delle due variabili sia molto piccolo, idealmente zero, cosicchè la traccia ricostruita punti esattamente al baricentro del *cluster* che contiene il deposito.
- HoE ,(H/E) : rappresenta la cosiddetta *frazione adronica* e si occupa di valutare il rapporto tra l'energia che viene depositata in HCAL in corrispondenza di un *supercluster* di ECAL e il deposito energetico in tale *supercluster*. È chiaro che per un elettrone questo rapporto debba tendere a zero dato che esso sciama e deposita tutta la sua energia all'interno del calorimetro elettromagnetico senza lasciare traccia all'interno del calorimetro adronico.

Nel prossimo capitolo andremo a vedere come valutare l'efficienza dei tagli effettuati sulle variabili che abbiamo introdotto.

Capitolo 3

Misure di efficienza con il metodo *Tag and Probe*

Un aspetto critico e delicato di qualsiasi analisi dati in un esperimento è sicuramente la stima dell'efficienza con cui si selezionano i dati. Generalmente l'approccio più utilizzato è quello di sfruttare una simulazione Monte Carlo per avere tale informazione, in questo lavoro di tesi vogliamo invece presentare un metodo alternativo che porta notevoli vantaggi: il *Tag and Probe*. Questo metodo, a differenza del Monte Carlo, permette di ottenere informazioni sull'efficienza direttamente dal campione di dati, inoltre elimina tutti gli errori sistematici propri di una simulazione, dovuti ad imperfezioni nella ricostruzione virtuale dell'esperimento. Il *Tag and Probe* [8] richiede l'utilizzo di una risonanza nota e semplice da analizzare, nel nostro caso un bosone Z^0 , per selezionare le particelle desiderate e misurare l'efficienza di un particolare pacchetto di selezioni su di esse. Vogliamo andare a cercare eventi di segnale che contengono un bosone Z, a tale scopo si cerca all'interno di ogni singolo evento l'elettrone con p_T più alto e al contempo in grado di soddisfare un pacchetto molto severo di tagli. Questo elettrone evidentemente è candidato ad essere uno dei due che formano la coppia leptonica in cui decade Z, e prende il nome di *Tag*. Con questo procedimento abbiamo dunque scartato eventi che sicuramente non contenevano un bosone Z e tuttavia nell'insieme di dati che abbiamo raccolto sono presenti anche i casi in cui esiste un elettrone candidato *Tag*, che soddisfa ai tagli richiesti ed ha il massimo p_T , ma in cui non c'è nessuno Z. Dunque il passo successivo è andare a cercare in questi eventi, dove si è trovato un candidato *Tag*, il secondo elettrone, sempre che ci sia, con più alto p_T , che prende il nome di *Probe*. Una volta fatto ciò si effettua una divisione degli eventi in due categorie:

passing probes : eventi in cui il *probe* soddisfa al pacchetto di selezioni di cui si intende valutare l'efficienza

failing probes : eventi in cui il *probe* non soddisfa al pacchetto di selezioni

Detto ciò possiamo dunque ora andare a definire l'*efficienza*:

$$\epsilon = \frac{N_{\text{passing}}}{N_{\text{passed}} + N_{\text{failing}}} \quad (3.1)$$

Innanzitutto occorre tenere ben presente che una volta ottenuti gli eventi divisi tra quelli contenenti *passing probes* e quelli contenenti *failing probes* si ottengono dei campioni che presentano un certo grado di rumore di fondo. Per poter ottenere informazioni fisiche accurate

dai dati occorre andare a quantificare il *rumore*, perciò risulta necessaria una procedura di *fit*. Si procede individuando una funzione analitica, con più parametri, in grado di riprodurre al meglio l'andamento del campione, in genere si scelgono due funzioni, una per il fondo e una per il segnale e se ne definisce una nuova, somma delle due. Con quest'ultima si va ad eseguire il *fit*, ovvero si cercano i valori dei parametri in grado di fornire la migliore descrizione dei dati a disposizione. Una volta ottenuto un buon risultato si vanno a cercare i parametri della funzione che riguardano la parte di segnale e usando tali parametri si disegna solo la quest'ultima. Idealmente questo dovrebbe rappresentare quello che si otterrebbe se non ci fosse rumore di fondo, ed è da questa funzione che prendiamo le informazioni per effettuare le nostre misure di efficienza.

3.1 Fit dei dati

Come detto nella procedura di *fit* occorre definire una funzione analitica e parametrica in grado di riprodurre al meglio l'andamento dei dati. La funzione complessiva, abbiamo già detto, è composta da una parte relativa al fondo e da una parte che riproduce il segnale. Nel presente lavoro di tesi il segnale è rappresentato dal picco di massa invariante della coppia leptonica in cui decade il bosone Z. Per il fondo, o *background* la scelta della funzione che lo riproducesse è stata piuttosto semplice, abbiamo optato per un esponenziale negativo che ci sembrava più flessibile della semplice retta a pendenza negativa:

$$\text{Background} = e^{-x}$$

Per il segnale invece è stato fatto un piccolo studio su tre diverse opzioni al fine di compiere la scelta ottimale:

$$\text{Gaussiana} = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{\frac{(x-\bar{x})^2}{2 \cdot \sigma^2}}$$

$$\text{Crystal Ball} = N \begin{cases} \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{\frac{(x-\bar{x})^2}{2 \cdot \sigma^2}} & \text{se } \frac{x-\bar{x}}{\sigma} > -\alpha \\ A \cdot (B - \frac{x-\bar{x}}{\sigma})^{-n} & \text{se } \frac{x-\bar{x}}{\sigma} \leq -\alpha \end{cases}$$

- A = $\left(\frac{n}{|\alpha|}\right)^n \cdot e^{-\frac{|\alpha|^2}{2}}$
- B = $\frac{n}{|\alpha|} - |\alpha|$
- N = normalizzazione

$$\text{Crujiff} = \begin{cases} e^{\frac{(x-\bar{x})^2}{\sigma_L^2}} & \text{se } x \leq \bar{x} \\ e^{\frac{(x-\bar{x})^2}{\sigma_R^2}} & \text{se } x > \bar{x} \end{cases}$$

In questa tesi il metodo *Tag and Probe* è stato studiato sullo standard di selezioni WP80. Quello che abbiamo fatto è stato applicare lo standard scelto ad entrambe le categorie di elettroni, il *Tag* ed il *Probe*, così facendo è chiaro che ci aspettiamo di ottenere un'efficienza di selezione che si aggira attorno all'80%. Per ottenere le informazioni desiderate abbiamo effettuato separatamente il *fit* sui dati relativi ai *Probes* che avevano soddisfatto le selezioni e quelli che

invece erano stati eliminati dai tagli. Una volta ottenuti i due fit con una qualità soddisfacente abbiamo estratto i parametri di interesse dalla funzioni di fit, che come già detto sono composte da fondo e segnale, per poter disegnare solo la funzione di segnale. Il passo successivo è stato integrarle in modo da ottenere il numero di eventi presenti nei picchi dei segnali, dopodichè usando la definizione che abbiamo già dato abbiamo valutato l'efficienza. Questa procedura è stata applicata in primo luogo per verificare quale tra le tre candidate fosse la funzione che meglio riproduceva il picco di massa invariante, in figura 3.1 vengono proposti i *plot* per i tre diversi casi.

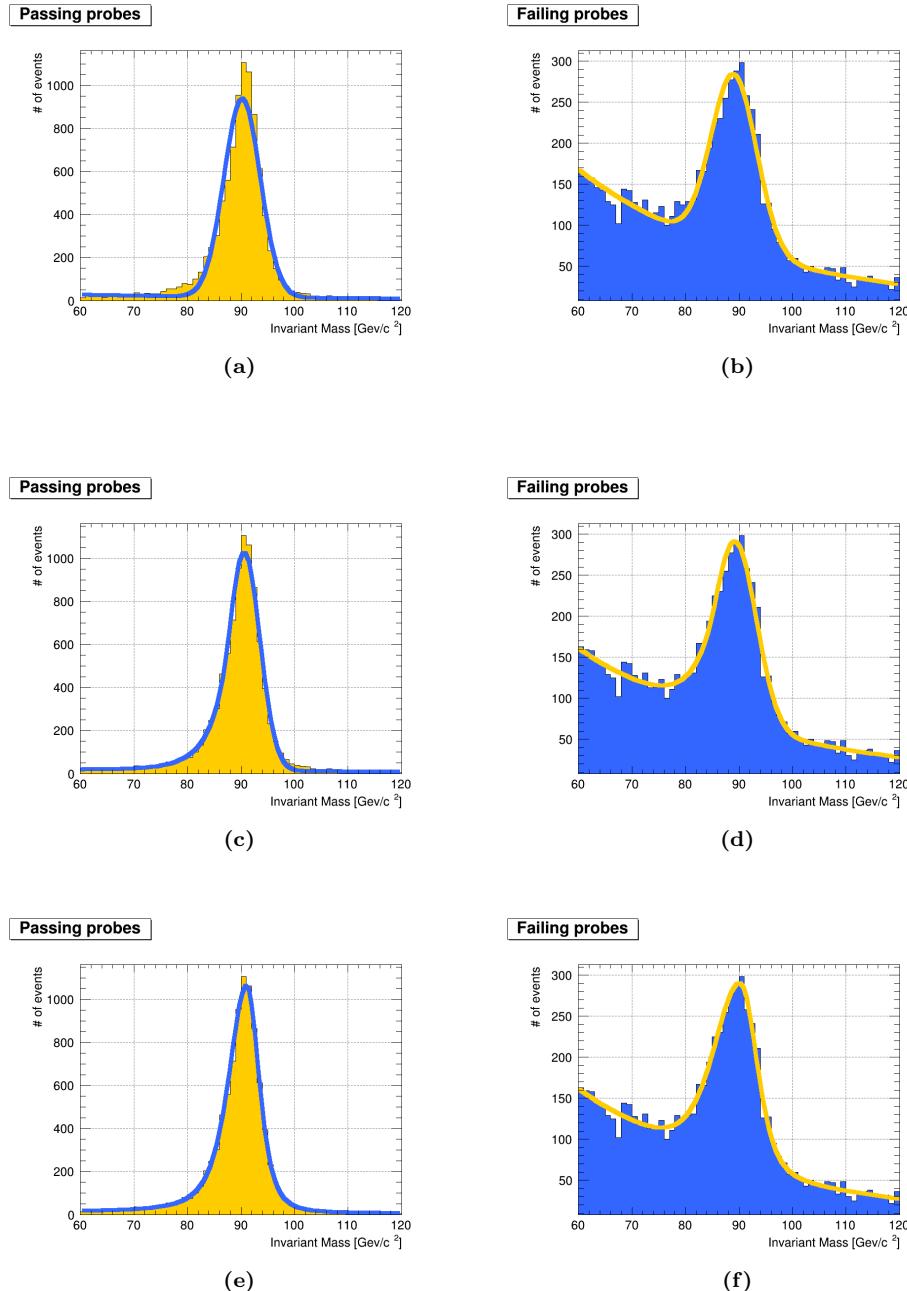


Figura 3.1: Grafici relativi allo studio effettuato con le tre funzioni di *fit*, in figura (a) e (b) la *Gaussian*, in figura (c) e (d) la *Crystal Ball*, in figura (e) ed (f) la *Crujiff*.

	Funzione	Efficienza	Errore Stimato	Errore Statistico
Gaussiana	0.775	4×10^{-5}	5×10^{-6}	
Crystalball	0.776	4×10^{-5}	5×10^{-6}	
Crujiff	0.788	2×10^{-4}	4×10^{-6}	

Tabella 3.1: Elenco dei valori di efficienza e degli errori, stimati e statistici, per le tre differenti funzioni: *Gaussiana*, *Crystal Ball* e *Crujiff*.

Dai valori ottenuti per le efficienze, tabella 3.1, è possibile indicare la *Crujiff* come la funzione che meglio descrive il nostro campione, tuttavia anche l'analisi grafica porta chiaramente alla stessa conclusione. Se andiamo ad osservare i grafici in figura 3.1 concentrando sui quelli relativi agli elettroni che hanno passato le selezioni (*Passing probes*) notiamo le funzioni *Gaussiana* e *Crystal Ball* non riescono a riprodurre al meglio il *dataset*. La funzione *Gaussiana* a causa della sua simmetria non riesce a tener conto della coda sinistra nella distribuzione del campione, inoltre non è in grado di riprodurre bene il picco di massa invariante. La *Crystal Ball* è una funzione asimmetrica, riproduce bene la coda sinistra della distribuzione ed il picco ma non è in grado di tener conto della coda destra. D'ora in poi dunque lavoreremo con la funzione *Crujiff* per effettuare le analisi di efficienza.

3.2 Studi di efficienza

Ora andremo ad illustrare alcuni dettagli tecnici rilevanti nello studio effettuato. Ogniqualvolta si effettua una misura occorre essere in grado di attribuire al valore ottenuto un errore in modo da avere informazioni sulla precisione della misurazione. Nel presente lavoro alle misure di efficienza sono stati associati due diversi errori, uno stimato e l'altro statistico. Abbiamo preferito seguire contemporaneamente due strade ai fini di avere una visione più ampia sulla problematica. L'errore da associare all'efficienza si ottiene in entrambi i casi propagando gli errori delle variabili presenti nella relazione 3.1 che fornisce l'efficienza ϵ . Nel nostro lavoro abbiamo trascurato nella propagazione il contributo dato dalle covarianze. Questo approccio trascura le correlazioni e massimizza l'errore ed è dunque piuttosto conservativo. Di seguito la relazione utilizzata:

$$\text{Errore} = \left| \frac{\partial \epsilon}{\partial N_{passing}} \right|^2 \sigma_{N_{passing}}^2 + \left| \frac{\partial \epsilon}{\partial N_{failing}} \right|^2 \sigma_{N_{failing}}^2 \quad (3.2)$$

Ciò che distingue le due categorie d'errore è il diverso modo di ottenere $\sigma_{N_{passing}}^2$ e $\sigma_{N_{failing}}^2$ che vanno associati alle variabili in gioco nell'efficienza. Nel caso dell'errore statistico le due quantità sono ottenute mediante le seguenti:

$$\sigma_{N_{passing}}^2 = \sqrt{N_{passing}} \quad (3.3)$$

$$\sigma_{N_{failing}}^2 = \sqrt{N_{failing}} \quad (3.4)$$

Per l'errore stimato invece abbiamo usato uno stratagemma piuttosto semplice, siamo andati a disegnare la funzione di segnale spostando i parametri prima al loro limite superiore e poi a quello inferiore. In questo modo abbiamo ottenuto fondamentalmente un *range* entro cui l'area della funzione di segnale, ovvero il numero di eventi di segnale, può variare. L'errore associato altro non è che la metà dell'intervallo individuato.

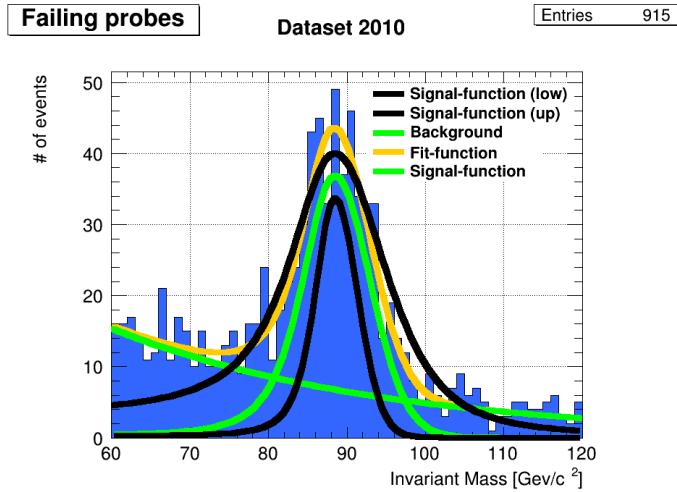


Figura 3.2: Grafico che mette in evidenza, in nero, le funzioni usate per la stima dell'errore, sono presenti anche le funzioni di segnale, di fondo e di fit.

$$\sigma_{N_{passing}}^2 = \frac{N_{passing-up} - N_{passing-down}}{2} \quad (3.5)$$

$$\sigma_{N_{failing}}^2 = \frac{N_{failing-up} - N_{failing-down}}{2} \quad (3.6)$$

Per chiarire meglio questo metodo proponiamo in figura 3.2 a un grafico in cui sono state disegnate le funzioni adottate per la stima dell'errore assieme a quelle relative al fondo al segnale e come al solito alla funzione di fit. Sottolineamo che d'ora in poi l'unica funzione che verrà presentata nei *plot* sarà quella di fit.

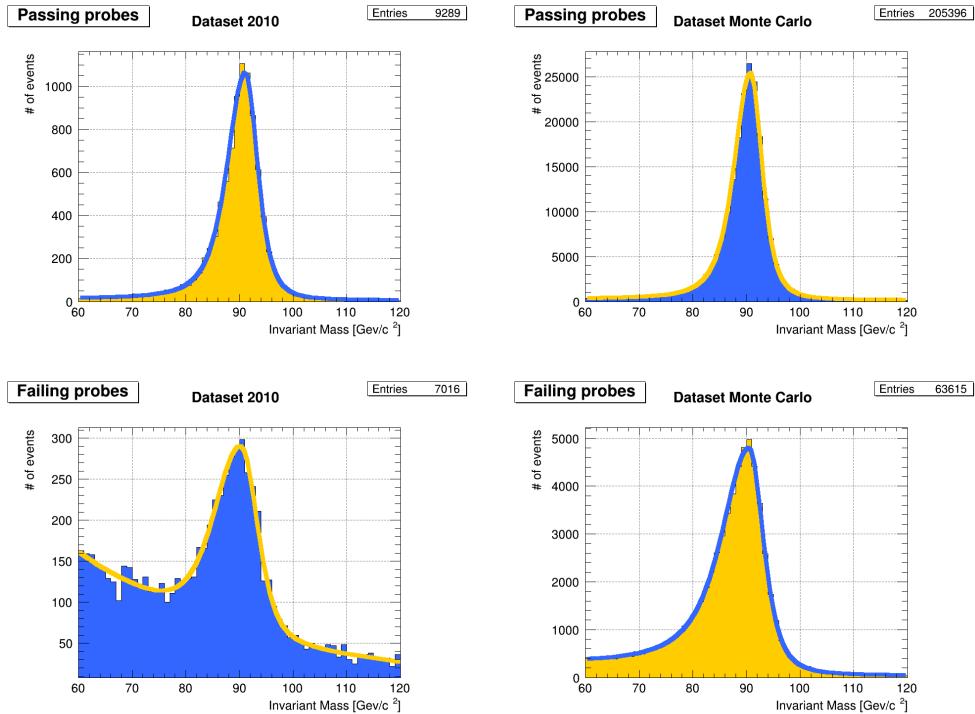
3.2.1 Studi specifici di efficienza

Ci occupiamo ora di presentare le analisi specifiche eseguite sull'efficienza, infatti oltre alla sua misurazione in questo lavoro ci siamo occupati di valutare come essa dipenda in modo specifico da alcune variabili. Nell'ambito di questa tesi abbiamo deciso di effettuare in parallelo lo studio sul nostro campione di dati e su un *dataset* generato con simulazione Monte Carlo. Questo accorgimento ci ha permesso, durante tutto lo studio, di avere un punto di riferimento in grado di avvalorare i nostri risultati e di indicarci eventuali errori commessi nelle analisi.

Efficienza sul campione inclusivo La prima misura è stata effettuata sul campione di dati inclusivo, ovvero considerando tutto il campione senza restrizione su alcuna variabile. Questa misura generale serve per ottenere l'efficienza vera e propria del nostro pacchetto di selezioni, il WP80, inoltre fornisce il valore di riferimento da utilizzare come linea guida nelle successive analisi. Come detto abbiamo sempre effettuato le nostre misure in parallelo sul campione di dati reale e su quello generato dal Monte Carlo proponiamo in figura 3.3 i grafici.

Dataset	Efficienza	Errore Stimato	Errore Statistico
Dataset 2010	0.788	2×10^{-4}	4×10^{-6}
Dataset Monte Carlo	0.766	6×10^{-6}	2×10^{-7}

Tabella 3.2: Valori di efficienza con errori stimati e statistici per il campione di dati inclusivo.

Figura 3.3: Grafici del *fit* sul dataset inclusivo.

Il risultato ottenuto non ci sorprende, siamo vicini al valore previsto dai tagli entro un intervallo inferiore al 5%. Il fatto che questo pacchetto di selezioni non restituisca un valore esatto di 0.8 è confermato dalla misura effettuata sul campione generato con simulazione Monte Carlo.

L'efficienza in funzione del momento trasverso del Probe La prima variabile su cui siamo andati a lavorare è il momento trasverso p_T . Dal momento che le particelle di cui ci interessiamo sono elettroni e positroni, dobbiamo ricordarci che esse vengono rilevate da *tracker* e calorimetro elettromagnetico. Abbiamo già accennato, nella descrizione di ECAL, al fatto che più energetica è la particella che viene rivelata è più alta è la precisione con cui essa è misurata. In aggiunta a questo occorre dire che l'elettrone (positrone) all'interno di CMS si muove su una traiettoria a spirale a causa del potente campo magnetico generato dal magnete solenoidale superconduttore. Come tutte le particelle cariche che viaggiano su una traiettoria non rettilinea l'elettrone (positrone) è soggetto a *bremsstrahlung*, ovvero emette radiazione. Quando si va a ricostruire l'energia del leptone si tiene chiaramente conto del fatto che esso ha perso energia per emissione di fotoni. Il problema sta nel fatto che più energetico è l'elettrone, ovvero più è alto il suo p_T , e più risulta accurato il processo che associa i vari fotoni agli elettroni da cui provengono, cerchiamo ora di chiarirne la ragione. Leptoni altamente energetici vengono curvati poco e dal momento che la radiazione di fotoni avviene lungo la tangente alla traiettoria nel momento dell'emissione ci aspettiamo di trovare l'elettrone e i fotoni da esso prodotti non troppo distanti tra loro. Questa configurazione rende molto accurato il processo di associazione dei fotoni all'elettrone da cui provengono. Se invece l'elettrone presenta un basso p_T la sua traiettoria sarà molto curva e perciò i fotoni che ha emesso verranno rivelati in genere lontano da esso distribuiti su un ampio intervallo di spazio. In tale configurazione invece l'accoppiamento dei fotoni con l'elettrone di provenienza diventa molto più delicato risultando più impreciso. Sulla base di quanto appena detto una particella con momento trasverso (e quindi energia) maggiore, contribuirà ad aumentare il valore dell'efficienza. Abbiamo diviso il nostro campione in tre intervalli di p_T e abbiamo ottenuto i risultati proposti nelle figure 3.4, 3.5 e 3.6.

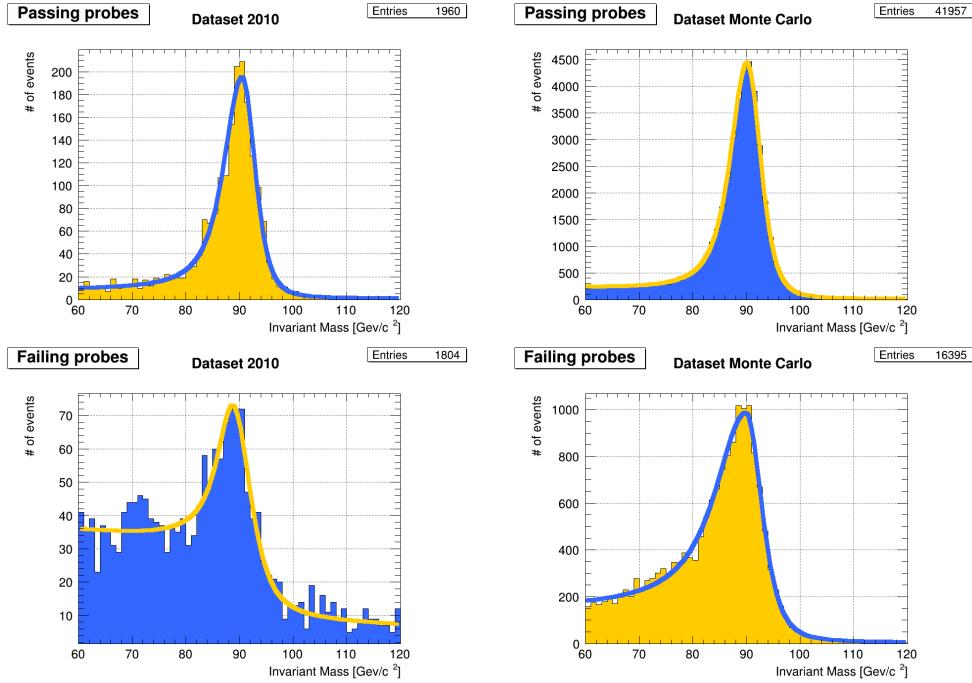


Figura 3.4: Grafici relativi al primo intervallo di momento trasverso: $20 \text{ GeV} < p_T < 30 \text{ GeV}$.

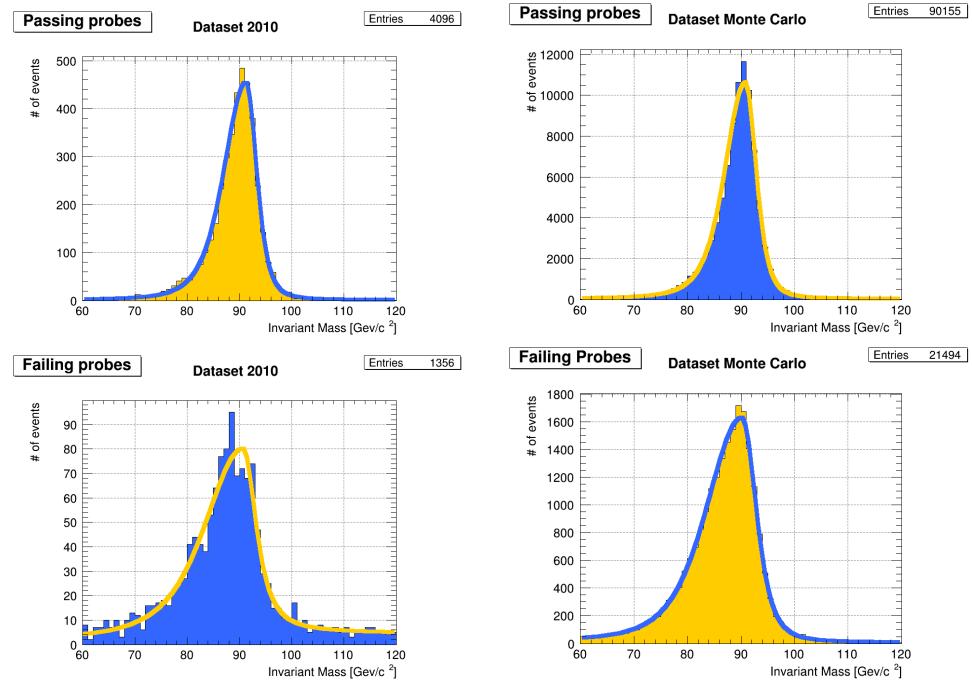


Figura 3.5: Grafici relativi al secondo intervallo di momento trasverso: $30 \text{ GeV} < p_T < 40 \text{ GeV}$.

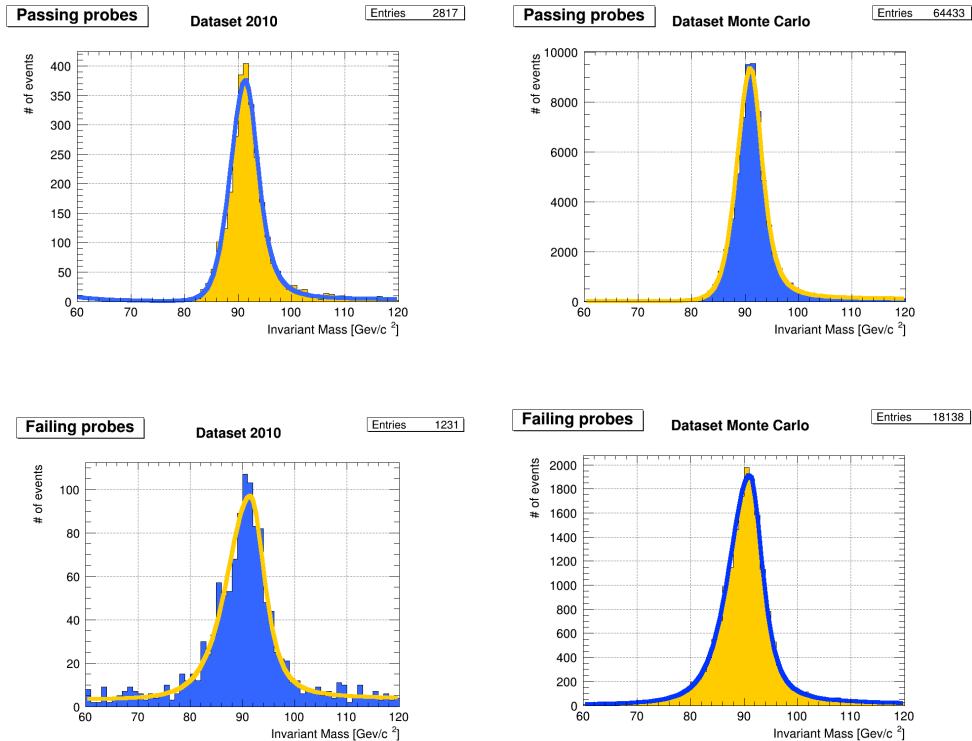


Figura 3.6: Grafici relativi all'ultimo intervallo di momento trasverso: $40 \text{ GeV} < p_T < \infty \text{ GeV}$.

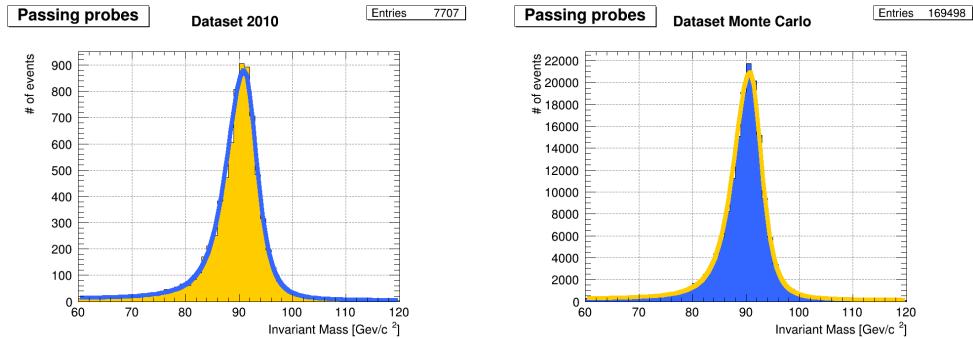
Dataset 2010				Dataset Monte Carlo			
Intervallo p_T	Eff.	Err. stim.	Err. stat.	Eff.	Err. stim.	Err. stat.	
20 - 30 GeV	0.614	1×10^{-2}	5×10^{-5}	0.714	5×10^{-5}	1×10^{-6}	
30 - 40 GeV	0.759	6×10^{-4}	1×10^{-5}	0.807	9×10^{-6}	3×10^{-7}	
40 - Inf Gev	0.727	1×10^{-3}	2×10^{-5}	0.772	2×10^{-5}	6×10^{-7}	

Tabella 3.3: Valori di efficienza ed errori per l'analisi dell'efficienza in funzione del momento trasverso.

Analizzando la tabella vediamo che come previsto il valore più basso per l'efficienza si ha nel range di momento trasverso più basso, per i motivi poc'anzi spiegati, risulta interessante invece che vi sia un'abbassamento nel terzo intervallo. Il fatto che questa diminuzione del valore dell'efficienza sia confermato anche dall'analisi effettuata sul campione generato con il Monte Carlo ci suggerisce che non si tratta di un problema relativo al campione di dati analizzato. Il fenomeno è spiegabile se si pensa alla nostra divisione dei *range* di momento. Quando ci mettiamo nel terzo intervallo andiamo a considerare, anche spingendoci all'estremo inferiore, elettroni altamente energetici, per i quali il meccanismo che associa i fotoni emessi per *bremssstrahlung* ai relativi elettroni è molto preciso. Questo fa sì che nella distribuzione del campione sparisca totalmente la coda sinistra. Tale coda è dovuta appunto a tutti quegli elettroni che principalmente a causa del loro basso p_T rendono poco preciso il meccanismo di associazione elettrone-fotoni di *bremssstrahlung*. Questa inferiore accuratezza fa sì che agli elettroni poco energetici venga attribuita un'energia inferiore a quella reale. Una minore energia associata ai leptoni fa in modo

modo che essi diano origine ad un valore di massa invariante inferiore al valore atteso. In questo modo il picco risulta leggermente asimmetrico poichè c'è comunque una componente di elettroni che producono un valore di massa invariante maggiore di quella attesa. Il fatto che la coda sinistra sia sparita fa sì che quella destra diventi importante e questo produce un alterazione dei parametri nel fit. Una distribuzione di massa invariante con una coda di eventi ad alti valori causa l'inversione del segno, dunque della pendenza, dell'esponente nella funzione che riproduce il rumore di fondo. Il rumore per quanto detto sull'efficienza del calorimetro elettromagnetico e sul fenomeno di *bremssstrahlung* dovrebbe essere importante per bassi valori di massa invariante, chiaramente causati dagli elettroni meno energetici. Per tale ragione la funzione che si occupa di riprodurre il fondo è un esponenziale decrescente, nel caso del terzo intervallo come detto tende però a diventare un esponenziale crescente. Per far sì che ciò non accada, in quanto un fondo crescente non avrebbe senso fisico, l'unica cosa da fare è imporre che l'esponente sia sempre negativo. Questa anomalia nei parametri causata dalla forzatura imposta dalla selezione dei *range* di momento va ad influire sulla capacità della funzione di *fit* di riprodurre al meglio l'andamento del campione. Questo fatto causa una diminuzione del valore dell'efficienza che risulta comunque tollerabile tenuto conto dei vincoli imposti alla funzione durante il *fit*.

L'efficienza in funzione del numero di getti adronici associati al *Probe* In questa sezione andremo ad illustrare lo studio fatto sul comportamento dell'efficienza al variare del numero di getti adronici associati all'elettrone che funge da *Probe*. La presenza di tali getti adronici influenza la capacità di rivelare con precisione la presenza di un elettrone, fondamentalmente rendono più difficile il riconoscimento del leptone da parte del calorimetro. Ci aspettiamo quindi che l'efficienza sia una funzione decrescente del numero di getti adronici associati al *Probe*. Avremmo voluto dividere la distribuzione della variabile che regola il numero di getti in tre intervalli. Ciò non è stato possibile in quanto nel campione a disposizione non c'era sufficiente statistica per ottenere risultati significativi nell'ultimo intervallo, quello con il maggior numero di getti associati. Ci siamo accontentati di suddividere il range della variabile in due parti: una relativa a *Probe* a cui non era associato alcun getto adronico ed una relativa a *Probe* a cui era associato almeno un getto. In figura 3.7 e 3.8 proponiamo i risultati ottenuti.



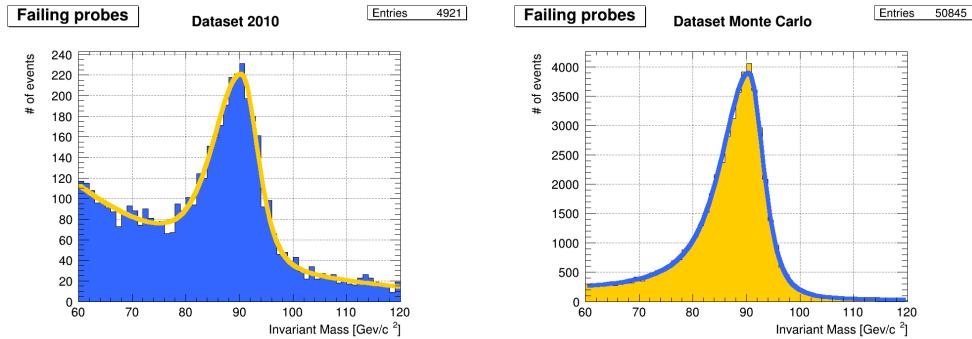


Figura 3.7: Grafici ottenuti in assenza di getti adronici associati al *Probe*.

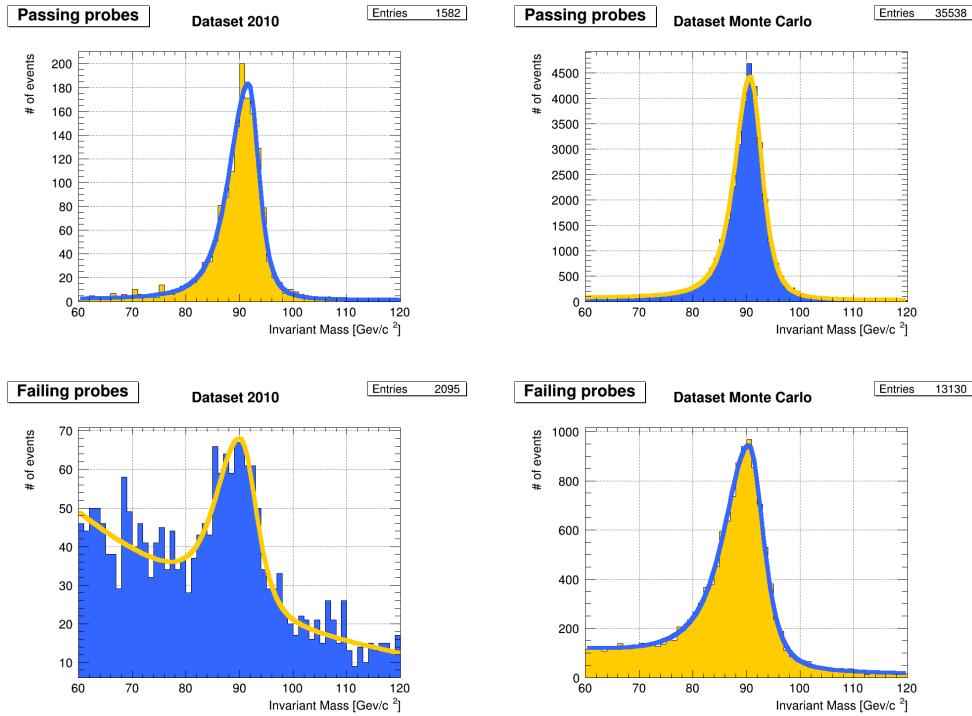


Figura 3.8: Grafici ottenuti in presenza di getti adronici.

Dataset 2010				Dataset Monte Carlo			
Numero getti	Eff.	Err. stim.	Err. stat.	Eff.	Err. stim.	Err. stat.	
$n_{Jets} = 0$	0.790	2.3×10^{-4}	4×10^{-6}	0.769	5×10^{-6}	2×10^{-7}	
$n_{Jets} > 0$	0.770	2×10^{-3}	3×10^{-5}	0.753	7×10^{-5}	1×10^{-6}	

Tabella 3.4: Valori di efficienza ed errori ottenuti nello studio effettuato sulla dipendenza dell'efficienza dal numero di getti adronici associati al *Probe*.

Come ci si aspetta si ottiene un valore di efficienza maggiore nel caso di getti adronici assenti. Anche in questo caso l'andamento è confermato dal Monte Carlo, infatti per entrambi i campioni c'è un calo di efficienza attorno al 3-4% nel passaggio da $n_{Jets} = 0$ a $n_{Jets} > 0$.

L'efficienza in funzione della *pseudorapidità* del *Probe* Come ultima analisi presentiamo ora quella relativa all'andamento dell'efficienza in funzione della *pseudorapidità* η , ovvero della posizione del *Probe* all'interno del calorimetro elettromagnetico. Come sappiamo ECAL come tutti i rivelatori di CMS è composto da due macro-componenti: il *barrel* e l'*endcap*. Nel primo vengono rivelati elettroni con alto p_T dal momento che per rimanere all'interno di esso le particelle devono avere un angolo significativo rispetto alla linea di fascio. Questa condizione permette agli elettroni di spendere la maggior parte della loro energia in momento trasverso e come abbiamo già più volte detto un alto p_T aumenta l'efficienza. Al contrario nell'*endcap* gli elettroni hanno basso momento trasverso poiché che la loro energia viene usata principalmente per il moto longitudinale. In aggiunta a questo data la configurazione spaziale, un elettrone che viene rivelato nell'*endcap* oltre ad attraversare gli strati di rivelatori del *tracker* ad un angolo minore, rendendo la tracciatura più imprecisa, deve attraversare molto più materiale. Dal momento che il leptone che arriva nell'*endcap* ha percorso più strada è più probabile che esso si sia formato per conversione di fotoni lungo il cammino. Chiaramente i tagli dovrebbero eliminare elettroni generati in tale processo ma questi non garantiscono un efficienza assoluta. La conseguenza di ciò è che ci si trova ad avere più rumore di fondo quando gli elettroni vengono rivelati nell'*endcap*. Si potrebbe effettuare un analisi per intervalli arbitrari di η ma dato che esiste già una divisione sensata suggerita dalla struttura del calorimetro come detto poco fa, la scelta più ragionevole è quelle di andare a considerare gli intervalli di *pseudorapidità* che individuano le zone di competenza di *barrel* ed *endcap*. Quello che vorremmo ottenere dunque è un valore di efficienza maggiore per i *Probe* rivelati nel *barrel* rispetto a quelli misurati nei due *endcap*. Nelle figure 3.9 e 3.10 illustriamo ciò che abbiamo ottenuto.

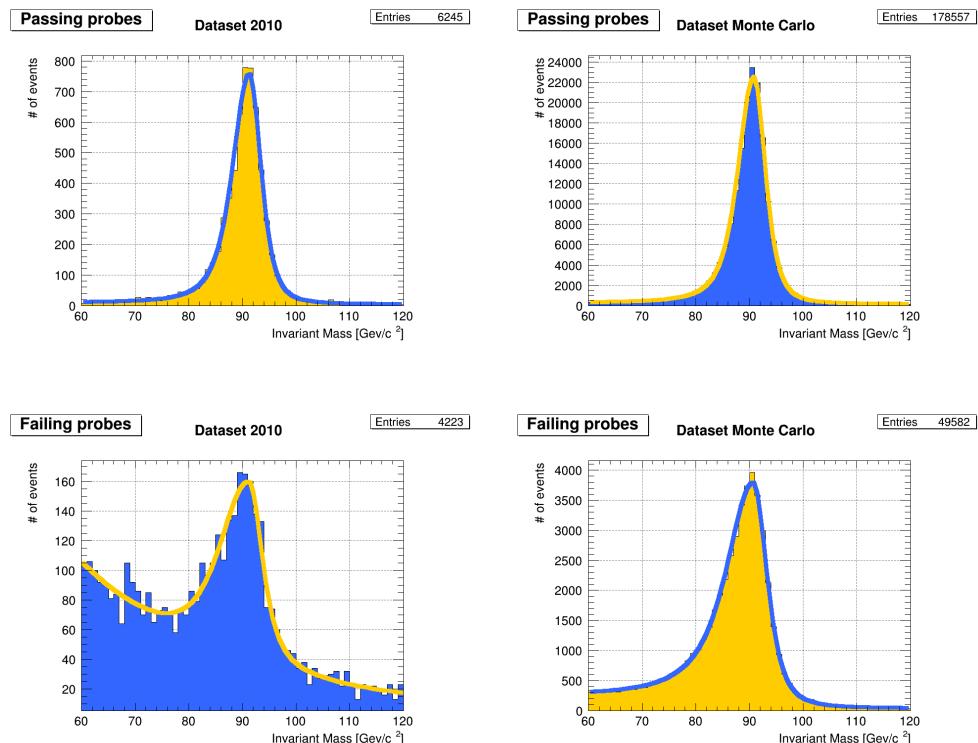


Figura 3.9: Grafici relativi all'analisi dell'efficienza per *probes* rilevati nel *barrel* ovvero avenuti $\eta < 1.4442$.

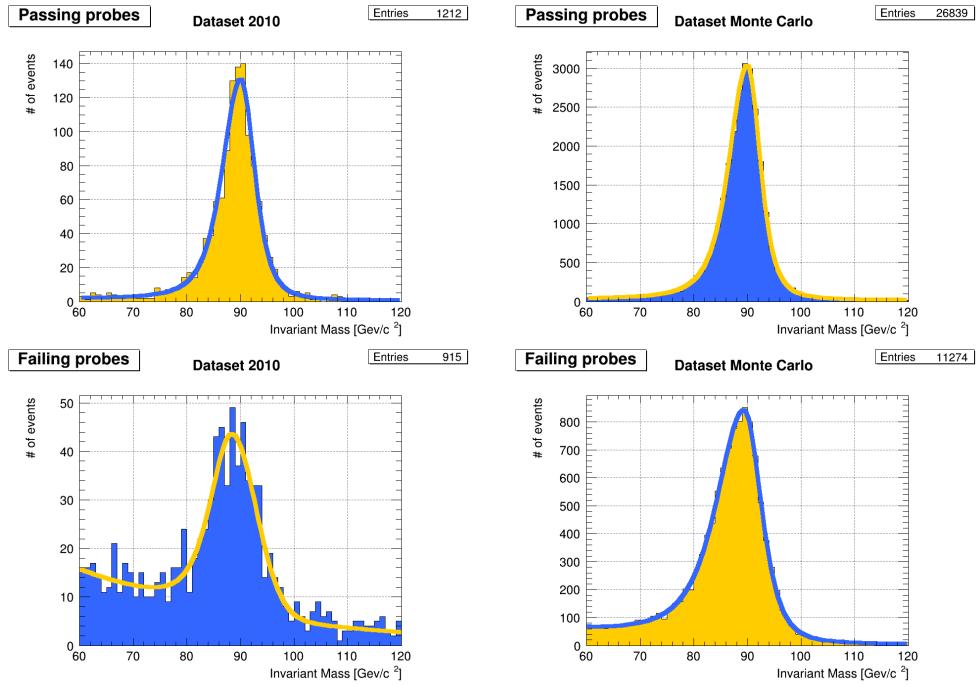


Figura 3.10: Grafici relativi all’analisi dell’efficienza per *probes* rilevati nell’*endcap* ovvero aventi $1.566 < \eta < 2.5$.

	Dataset 2010				Dataset Monte Carlo			
η	Eff.	Err. stim.	Err. stat.		Eff.	Err. stim.	Err. stat.	
$\eta < 1.4442$	0.779	1×10^{-3}	6×10^{-6}		0.783	6×10^{-6}	2×10^{-7}	
$1.566 < \eta < 2.5$	0.735	3×10^{-3}	4×10^{-5}		0.715	1×10^{-4}	2×10^{-6}	

Tabella 3.5: Valori di eficienza ed errori ottenuti nello studio dell’efficienza in funzione della variabile η .

In accordo con le previsioni l’efficienza è maggiore quando i *probes* vengono rilevati all’interno del *barrel*. Questo risultato è in accordo con il fatto che in tale zona del rivelatore le particelle hanno p_T più alto e il meccanismo di riconoscimento degli elettroni è più accurato.

Conclusioni

In questo lavoro di tesi abbiamo dimostrato come il metodo *Tag and Probe* sia un valido strumento di indagine fisica quando si tratta di misure di efficienza. Premettiamo che il campione di dati con cui abbiamo lavorato non conteneva un numero elevato di eventi. Per tale ragione gli studi differenziali effettuati sulle singole variabili sono stati forzatamente fatti su intervalli piuttosto ampi della variabile di interesse. Questo limite ha fatto sì che il numero di tali intervalli fosse piuttosto ridotto. In aggiunta a questo, occorre dire che lo standard di selezioni WP80 utilizzato da noi non è stato tarato sul nostro campione di dati. Dopo i *run* del 2010 effettuati ad LHC tutti gli standard di selezione sono stati rivisti e corretti per fornire con più accuratezza il valore di efficienza che nominalmente garantiscono.

Nonostante i limiti entro cui ci siamo dovuti muovere nel nostro lavoro, siamo comunque riusciti ad ottenere risultati consistenti con quanto previsto dalla teoria. Abbiamo dapprima visto come il nostro algoritmo fornisse il corretto valore di efficienza rispetto allo standard utilizzato. Come già detto il valore è stato ottenuto con un errore del 5% a causa della taratura non ottimale delle selezioni sul nostro campione. Gli studi sulle singole variabili hanno tutti prodotto i risultati attesi tranne apparentemente quello in funzione del momento trasverso p_T . Apparentemente poichè in realtà tale deviazione dal comportamento previsto rientra comunque nel margine di errore causato dai problemi, discussi nel capitolo relativo a questo studio, riguardanti la funzione di *fit*. Possiamo dunque affermare che il metodo *Tag and Probe* è senz'altro uno strumento efficace e flessibile per questo tipo di misura.

L'appuccio utilizzato nel nostro lavoro, ovvero l'utilizzo parallelo del *Tag and Probe* sul nostro campione e su quello simulato, permette anche di ottenere informazioni sul rapporto tra i valori ottenuti con l'uno e con l'altro. Nello specifico è possibile ottenere la costante, detta fattore di scala, che moltiplicata per il valore di efficienza prodotto con il campione Monte Carlo fornisce il valore vero prodotto dal *Tag and Probe* sul campione di dati reale.

Bibliografia

- [1] A.Schizzi, *Risoluzione spaziale del calorimetro elettromagnetico dell'esperimento CMS*, Tesi Triennale, July 2008, consultabile all' indirizzo http://www.infn.it/thesis/thesis_dettaglio.php?tid=2646
- [2] CERN public website, *Triggering and Data Acquisition*, visitabile all'indirizzo <http://cms.web.cern.ch/news/triggering-and-data-acquisition>
- [3] CMS, *The TriDAS Project, Technical Design Report Volume 1: The Trigger Systems*, December 2010 , consultabile all'indirizzo <http://cds.cern.ch/record/706847/files/cer-002248791.pdf>
- [4] C. Foudas, *The CMS Level-1 Trigger at LHC and Super-LHC*, Imperial College preprint, consultabile all'indirizzo <http://workspace.imperial.ac.uk/highenergyphysics/Public/preprints/08-19.pdf>.
- [5] S.Baffioni, *Electron Reconstruction at CMS*, CMS Note, February 2006, consultabile all'indirizzo http://hal.in2p3.fr/docs/00/14/04/84/PDF/NOTE2006_040.pdf
- [6] The CMS Trigger and Data Acquisition Group, *The CMS High Level Trigger*, CMS Record, November 2005, consultabile all' indirizzo <http://cds.cern.ch/record/919323/files/0512077.pdf?version=1>
- [7] W.Andrews, *Study of photon conversion rejection at CMS*, CMS Internal Note, July 2009, consultabile all'indirizzo <http://uaf-2.t2.ucsd.edu/~yanjuntu/conversion09.pdf>
- [8] A.Schizzi, *Studio della produzione associata di getti adronici e bosoni vettori Z e W in collisioni p-p nell'esperimento CMS*, Tesi Specialistica, July 2011, consultabile all' indirizzo <http://www.infn.it/thesis/PDF/getfile.php?filename=5837-Schizzi-specialistica.pdf>.