UNIVERSITY OF IOANNINA

DEPARTMENT OF PHYSICS

Study of multi-parameter adaptive simulations of new physics signal.

Author:

Nicholas Karatzenis

Supervisor:

John Strologas

October 2021

ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ

ΤΜΗΜΑ ΦΥΣΙΚΗΣ

Μελέτη πολύ-παραμετρικής και προσαρμοστικής προσομοίωσης σήματος νέας φυσικής.

Συγγραφέας:

Νικόλαος Καρατζένης

Επιβλέπων:

Ιωάννης Στρόλογγας

Οκτώβριος 2021

Examination Committee

John Strologas, Assistant Professor (chair)

Costas Fountas, Professor

Nikos Manthos, Professor

Contents

Abstract

In this thesis we show how a new multi-parameter integration algorithm can be used to successfully integrate not only the kinematic space of a theory but also its parameter space. The result of the adaptive integration is a generator that samples simultaneously both particle kinematics and theory parameters. Applications of this method include the efficient setting of new-physics limits in multi-parameter space, the optimization of signal selection (because of the direct connection of kinematics to parameters) and the determination of parameter space consistent with signal in case of discovery.

We demonstrate this method in the case of Z' boson production at the LHC with two free parameters: the mass of the boson and a factor that multiplies the Standard-Model coupling constants of its decay to leptons. We establish that the integration of both kinematics and parameters converges and can be used to determine cross sections for each point of parameter space. We conclude that with a simple Monte Carlo run we can determine a finer two-dimensional 95% confidence-level exclusion region of new physics or, in the case of discovery, we are able to determine the allowed parameter space.

Περίληψη

Η σωματιδιακή φυσική αποτελεί τον κλάδο της φυσικής που ασχολείται με την μελέτη της θεμελιακής δομής του κόσμου γύρω μας. Η καλύτερη θεωρία που έχει αναπτυχθεί για την κατανόηση αυτή είναι το Καθιερωμένο Πρότυπο. Το Καθιερωμένο Πρότυπο αποτελείται από θεωρίες βαθμίδας που αξιοποιούν τον φορμαλισμό κατά Lagrange ώστε να εξηγήσουν τα πεδία και τις αλληλεπιδράσεις που παρατηρούμε γύρω μας. Οι Λαγκραντζιανές της θεωρίας υπαγορεύονται από τις συμμετρίες στις οποίες υπακούουν οι αλληλεπιδράσεις (ηλεκτρομαγνητική, ισχυρή πυρηνική, ασθενής πυρηνική). Το Καθιερωμένο Πρότυπο εξηγεί (και έχει προβλέψει σε κάποιες περιπτώσεις) την ύπαρξη πολλών θεμελιωδών σωματιδίων τα οποία παρατηρούνται γύρω μας. Όμως ακόμη υπάρχουν ερωτήματα που δεν έχουν απαντηθεί από την θεωρία καθώς και προβλέψεις που δεν έχουν επικυρωθεί.

Για αυτό το λόγω γίνονται πειράματα με σκοπό την ανακάλυψη νέας φυσικής. Ο βασικός πειραματικός τρόπος είναι η μελέτη αλληλεπιδράσεων των θεμελιωδών σωματιδίων και εξέταση για κάποια συμπεριφορά διαφορετική από αυτήν που αναμένουμε σύμφωνα με το Καθιερωμένο Πρότυπο. Τα πιο μεγάλα πειράματα που μελετούν αυτές τις αλληλεπιδράσεις λαμβάνουν χώρα αυτή τη στιγμή στον Μεγάλο Αδρονικό Επιταχυντή (Large Hadron Collider, LHC) στο CERN στον οποίον έχουμε συγκρούσεις πρωτονίων με ενέργεια στο κέντρο μάζας της σύγκρουσης $\sqrt{s} = 13$ TeV.

Η κατανόηση των αλληλεπιδράσεων που λαμβάνουν χώρα κατά τη διάρκεια μιας σύγκρουσης είναι πολύπλοκο ζήτημα. Για αυτό το λόγο, χρησιμοποιούμε προσομοιώσεις των συγκρούσεων ώστε να αναπαραστήσουμε τα γεγονότα που τελικά παρατηρούμε στο πείραμα και να βγάλουμε συμπεράσματα. Οι προσομοιώσεις αυτές λειτουργούν ως γεννήτορες σωματιδίων ολοκληρώνοντας κατά Monte Carlo τον κινηματικό χώρο της θεωρίας (συνήθως για τον υπολογισμό της ενεργού διατομής) με χρήση δειγματοληψίας σπουδαιότητας ώστε να μας δώσει την παραγωγή σωματιδίων όπως περιμένουμε να τα παρατηρήσουμε στο πείραμα. Αυτό συμβαίνει επειδή μεγαλύτερη τιμή της ολοκληρωτέας ποσότητας σημαίνει μεγαλύτερη ενεργός διατομή, άρα πιο πιθανή σωματιδιακή διεργασία.

Σκοπός της εργασίας μας είναι να επεκτείνουμε αυτήν την ολοκλήρωση ώστε να συμπεριλαμβάνει πέρα από τον κινηματικό χώρο της θεωρίας και έναν παραμετρικό χώρο όπου μπορούμε παράλληλα να ελέγξουμε και τις ελεύθερες παραμέτρους της θεωρίας. Με αυτόν τον τρόπο μπορούμε να μελετήσουμε ολόκληρο τον παραμετρικό χώρο της θεωρίας με ένα μόνο αρχείο προσομοίωσης. Μας επιτρέπεται, έτσι, η άμεση σύνδεση της κινηματικής των σωματιδίων με τις τιμές των παραμέτρων νέων θεωριών. Κύριες εφαρμογές της μεθόδου συμπεριλαμβάνουν την εύκολη βελτιστοποίηση της κινηματικής επιλογής γεγονότων για την έρευνα θεωριών νέας φυσικής και τον αποδοτικό καθορισμό ορίων αποκλεισμού ή ανακάλυψης για τις θεωρίες αυτές.

Για να το επιτύχουμε αυτό θα σχεδιάσουμε πρώτα έναν γεννήτορα σωματιδίων για να μελετήσουμε την αλληλεπίδραση Drell-Yan σε συνθήκες του LHC και κατόπιν θα επεκτείνουμε την ολοκλήρωση ώστε να συμπεριλαμβάνει και ελεύθερες παραμέτρους της θεωρίας.

Η αλληλεπίδραση Drell-Yan προκύπτει από την σύγκρουση ενός quark και ενός antiquark τα οποία μας δίνουν ένα φωτόνιο ή ένα μποζόνιο Z τα οποία μας δίνουν ως τελικό προϊόν ένα ζεύγος

λεπτονίων. Μελετήσαμε τις συγκρούσεις όλων των ζευγών quark, εκτός από το ζεύγος top με anti-top καθώς διαθέτουν πολύ υψηλή μάζα. Επίσης εξετάσαμε μόνο την περίπτωση του ζεύγους ηλεκτρονίου-ποζιτρονίου ως τελικό προϊόν για λόγους απλούστευσης χωρίς να χάνεται η γενικότητα.

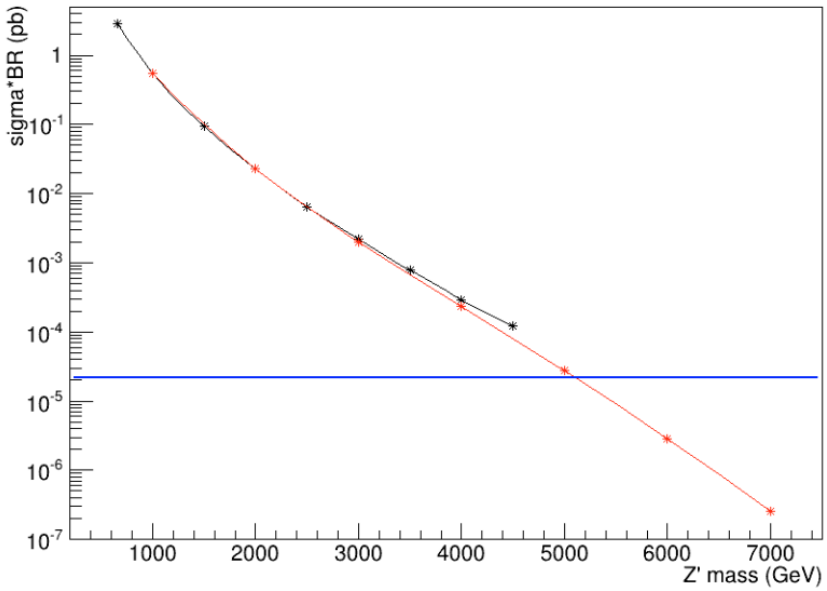
Η ενεργός διατομή της αλληλεπίδρασης δίνεται από την εξίσωση (Eq. 5 στο κείμενο):

$$\sigma = \int \sum_{q_i,q_j} f_i(x_i,\mu) f_j(x_j,\mu) \left|M(P_\iota, P_j, P_{out,k})\right|^2 \delta^4(P_\iota + P_j - \sum_k P_{out,k}) \prod_k \frac{dP_{out,k}}{E_k} \frac{dx_i dx_j}{E_i E_j},$$

την οποία υλοποιούμε στον κώδικα. Οι συναρτήσεις $f_{i,j}$ είναι τα PDFs (Particle Distribution Functions) που μας δίνουν την πιθανότητα το εκάστοτε quark ή antiquark $q_{i,j}$ εντός του πρωτονίου να έχει ποσοστό ορμής $x_{i,j}$. Τα PDFs προκύπτουν από προσαρμογές πειραματικών δεδομένων από πολλά πειράματα και εξαρτώνται από μια ενεργειακή κλίμακα μ. $P_{i,j}$ είναι οι ορμές των συγκρουόμενων quark και $P_{out,k}$ οι ορμές των εξερχόμενων λεπτονίων και $E_{i,j}$ οι αντίστοιχες ενέργειες. Το στοιχείο πίνακα *M* (matrix element) μας δίνει την πιθανότητα να συγκρουστούν δύο quarks και να παράξουν το μποζόνιο και μετά δύο λεπτόνια, με τις δεδομένες ορμές. Αυτόν τον υπολογισμό τον παίρνουμε σε πρώτο βαθμό προσέγγισης (leading order) από το Madgraph. Η *συνάρτηση δέλτα* εκφράζει την αρχή διατήρησης της ορμής και είναι επίσης ενσωματωμένη μέσα στο πρόγραμμα. Η ολοκλήρωση, όπως βλέπουμε, γίνεται πάνω στις κινηματικές μεταβλητές.

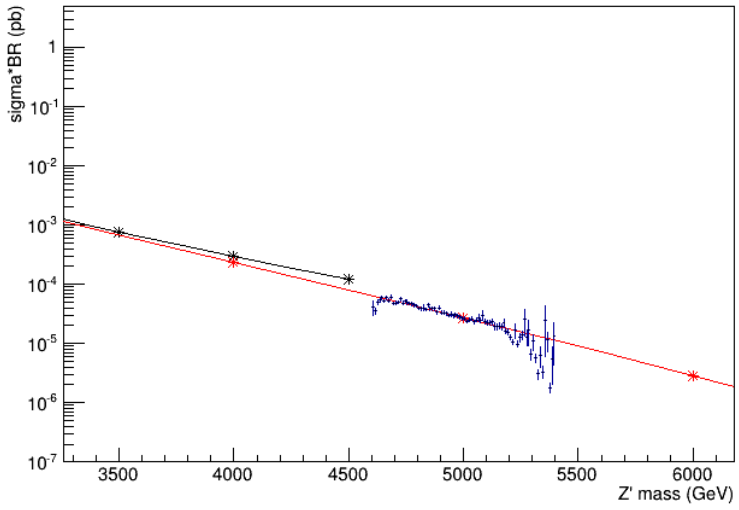
Η παραπάνω ολοκλήρωση γίνεται με τον αλγόριθμο VEGAS που χρησιμοποιεί δειγματοληψία σπουδαιότητας καθώς και stratified. Ενδείκνυται για πολυδιάστατες ολοκληρώσεις καθώς είναι δυναμικός, επαναληπτικός και οι υπολογιστικές του ανάγκες, ανάλογα με τις διαστάσεις ολοκλήρωσης, αυξάνονται γραμμικά και όχι εκθετικά. Η ολοκλήρωση πραγματοποιείται με χρήση τεχνικών Monte Carlo καθώς και με βελτιστοποίηση του πλέγματος στο οποίο λαμβάνει χώρα η ολοκλήρωση. Είναι ιδιαίτερα χρήσιμος σε μας γιατί έχουμε επιπλέον διαστάσεις ολοκλήρωσης λόγω της ταυτόχρονης ολοκλήρωσης και στον παραμετρικό χώρο της θεωρίας.

Το επόμενο βήμα είναι είναι να συμπεριλάβουμε στην ολοκλήρωση τις παραμέτρους θεωρίας νέας φυσικής. Προκειμένου να εφαρμόσουμε τη μέθοδο, μελετάμε την παραγωγή μποζονίου Z' μεταβάλλοντας δύο παραμέτρους της διεργασίας Drell-Yan: Τη μάζα του μποζονίου και έναν παράγοντα που πολλαπλασιάζει την σταθερά σύζευξης του μποζονίου με τα λεπτόνια στα οποία διασπάται. Το πρώτο μας βήμα είναι, κάνοντας χρήση του προγράμματος, να δώσουμε διαφορετικές τιμές στην μάζα (με τη σταθερά σύζευξης στην τιμή του Καθιερωμένου Προτύπου) και να κάνουμε ξεχωριστές ολοκληρώσεις για να υπολογίσουμε την αντίστοιχη προκύπτουσα ενεργό διατομή (μετά και από κάποιες επιπλέον διορθώσεις). Σχεδιάζουμε την γραφική παράσταση της ενεργού διατομής, που υπολογίσαμε, σε συνάρτηση με τη μάζα του μποζονίου, όπως φαίνεται στο Σχήμα Α. Στο ίδιο σχήμα παρουσιάζουμε την υπολογισμένη αναλυτικά ενεργό διατομή ως συνάρτηση της μάζας με σταθερές σύζευξης σύμφωνες με το Καθιερωμένο Πρότυπο και παρατηρούμε πολύ καλή συμφωνία, με βάση τη θεωρητική αβεβαιότητα. Η οριζόντια γραμμή του σχήματος αντιστοιχεί στο όριο αποκλεισμού ενεργών διατομών, σε περίπτωση που ανιχνεύσουμε μηδέν γεγονότα με τέλειο ανιχνευτή και χωρίς να αναμένουμε υπόβαθρο από το Καθιερωμένο Πρότυπο. Σε αυτήν την περίπτωση όλες οι ενεργές διατομές πάνω από την οριζόντια γραμμή απορρίπτονται με 95% confidence level (CL) που σημαίνει ότι απορρίπτουμε μάζες του Z' κάτω από 5.11 TeV.



Σχήμα Α : Σύγκριση μεταξύ της θεωρητικής ενεργού διατομής (μαύρα σημεία) με τα αποτελέσματα του προγράμματος μας (κόκκινα σημεία). Στο σχήμα συμπεριλαμβάνεται και το πάνω όριο (95% CL) των επιτρεπτών ενεργών διατομών για τη φωτεινότητα του LHC, υπό την υπόθεση τέλειου ανιχνευτή και μηδενικού υποβάθρου από το Καθιερωμένο Πρότυπο (μπλε γραμμή). [Figure 7 στο κυρίως κείμενο.]

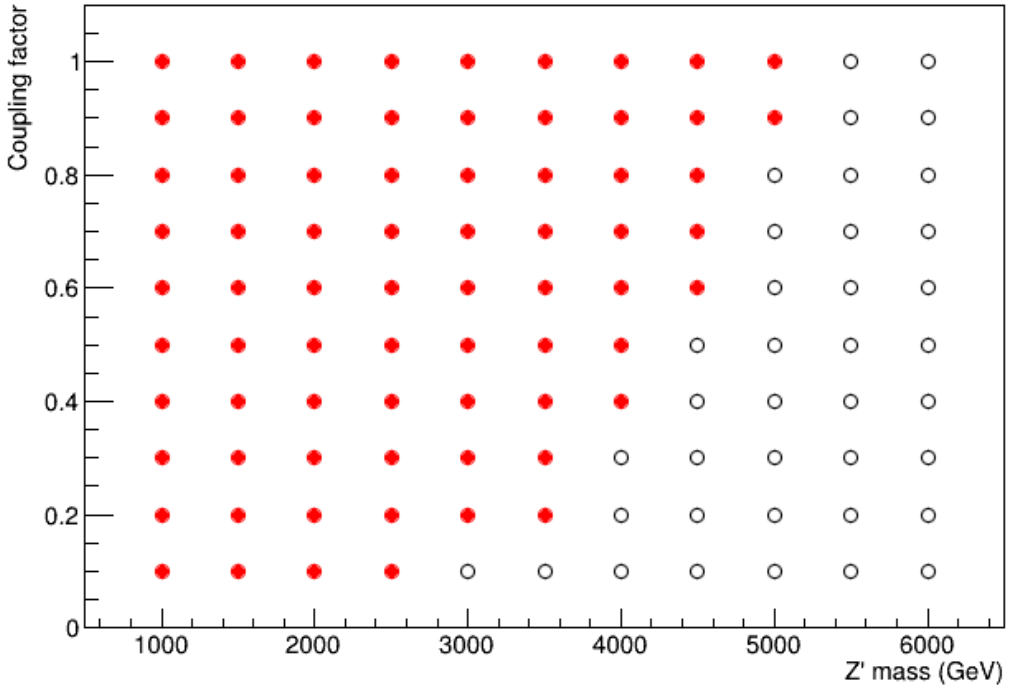
Από τη στιγμή που είμαστε σίγουροι πως υπολογίζουμε σωστές ενεργές διατομές μπορούμε να αφήσουμε την μάζα ως ελεύθερη παράμετρο, δηλαδή μεταβλητή ολοκλήρωσης (θέσαμε όρια ολοκλήρωσης μεταξύ 4600 GeV και 5400 GeV). Από την νέα ολοκλήρωση υπολογίζουμε τις τιμές της ενεργού διατομής και καταλήγουμε ξανά στο προηγούμενο διάγραμμα όμως αυτή τη φορά με μεγαλύτερη διακριτική ικανότητα (10 GeV), όπως αυτή φαίνεται στο Σχήμα Β.



Σχήμα Β : Η ενεργός διατομή για μάζα $M_{Z'}$ μεταξύ 4600 και 5400 GeV όπως προέκυψε από την ολοκλήρωση της παραμέτρου μάζας (μπλε σημεία), σε σύγκριση με τους ξεχωριστούς υπολογισμούς δεδομένης τιμής της μάζας (κόκκινα σημεία και γραμμή). [Figure 8 στο κυρίως κείμενο.]

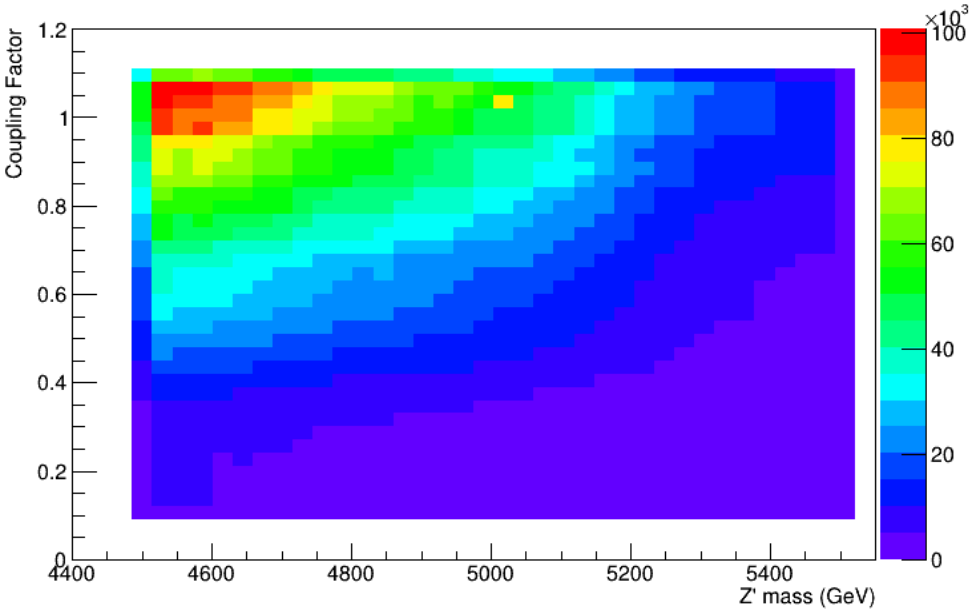
Παρατηρούμε κάποιον θόρυβο στις υψηλές μάζες ο οποίος διορθώνεται είτε με αύξηση των στατιστικών δεδομένων (περισσότερος χρόνος υπολογισμών) είτε χρησιμοποιώντας μικρότερο εύρος ολοκλήρωσης.

Στη συνέχεια αφήνουμε ως επιπλέον ελεύθερη παράμετρο τον παράγοντα σύζευξης. Ο παράγοντας αυτός για το καθιερωμένο πρότυπο είναι ίσος με τη μονάδα. Όπως πριν, υπολογίζουμε την ενεργό διατομή της διεργασίας Drell-Yan για συγκεκριμένες τιμές της μάζας αλλά και του συντελεστή σύζευξης και προκύπτει το δισδιάστατο διάγραμμα του Σχήματος Γ, στο οποίο κάθε σημείο αντιστοιχεί σε συνδυασμό της μάζας και του παράγοντα σύζευξης. Στο σχήμα σημειώνουμε με κόκκινο χρώμα τα σημεία που αποκλείονται με 95% CL, αν δεν παρατηρήσουμε δεδομένα, σε τέλειο ανιχνευτή και χωρίς υπόβαθρο. Ο σκοπός μας είναι να αντικαταστήσουμε αυτό το διάγραμμα αποκλεισμού με ένα που έχει μεγαλύτερη ακρίβεια και χωρίς την επανάληψη της προσομοίωσης για διαφορετικό συνδυασμό παραμέτρων.



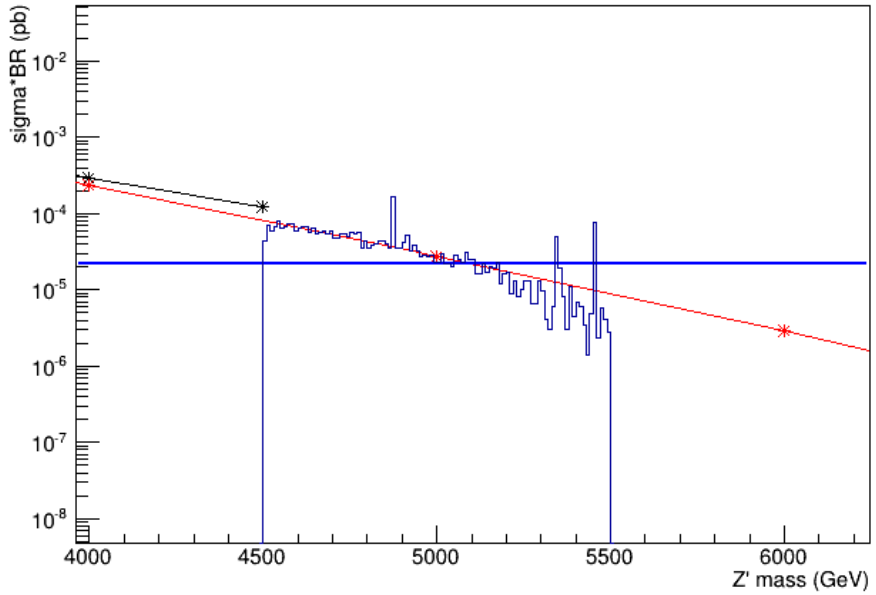
Σχήμα Γ : Επαναλαμβανόμενες ολοκληρώσεις Monte Carlo για δεδομένες τιμές παραμέτρων ώστε να διακρίνουμε την περιοχή αποκλεισμού. Με κόκκινο σημειώνουμε τα σημεία του παραμετρικού χώρου που αποκλείονται με 95% CL, αν δεν ανιχνεύσουμε γεγονότα και θεωρώντας τέλειο ανιχνευτή και μηδενικό υπόβαθρο. [Figure 10 στο κυρίως κείμενο.]

Στη συνέχεια, ομοίως με πριν, αφήνουμε ως ελεύθερες παραμέτρους την μάζα του Z’ και τον παράγοντα σύζευξης, οι οποίες γίνονται τώρα και οι δύο μεταβλητές ολοκλήρωσης. Αρχικά εξετάζουμε την λειτουργία του προγράμματος ως γεννήτορα γεγονότων στον παραμετρικό χώρο της θεωρίας πράγμα που επιβεβαιώνεται καθώς έχουμε μεγαλύτερη παραγωγή σωματιδίων για μικρές μάζες Z’ και μεγάλους συντελεστές σύζευξης, όπως φαίνεται στο Σχήμα Δ.



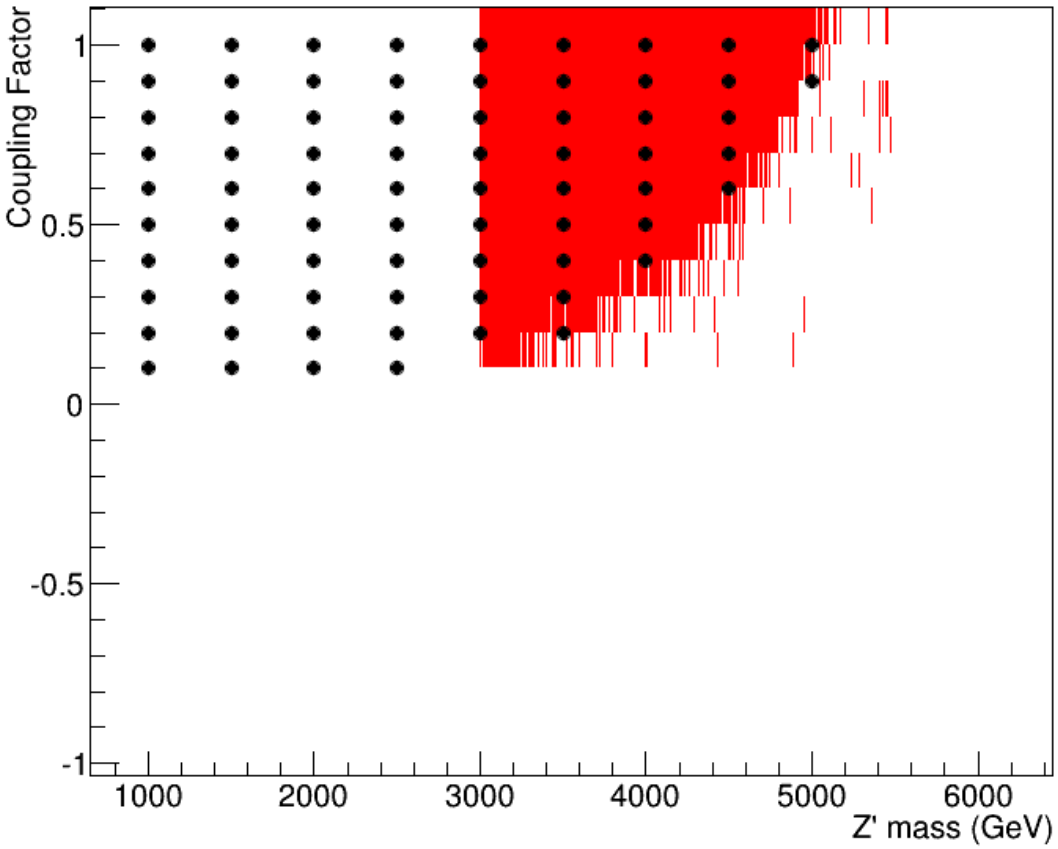
Σχήμα Δ : Η δειγματοληψία στις δύο (παραμετρικές) από τις έξι διαστάσεις ολοκλήρωσης με τη μάζα $M_{Z'}$ μεταξύ 4500 και 5500 GeV και παράγοντα σύζευξης μεταξύ 0.1 και 1.1. [Figure 11 στο κυρίως κείμενο.]

Επίσης, πριν παράγουμε το τελικό μας διάγραμμα αποκλεισμού, επιβεβαιώνουμε ξανά πως συνεχίζουμε να υπολογίζουμε σωστές ενεργές διατομές. Για να το πετύχουμε αυτό θέτουμε τον συντελεστή σύζευξης μεταξύ 0.95 και 1.05 προσεγγίζοντας δηλαδή τον παράγοντα του καθιερωμένου προτύπου και συγκρίνουμε ξανά με τη γραφική παράσταση του Σχήματος Α όπου για ακόμη μια φορά παρατηρούμε πολύ καλή συμφωνία.



Σχήμα Ε : Η ενεργός διατομή όπως αυτή υπολογίζεται από το πρόγραμμα μας ως συνάρτηση της μάζας $M_{Z'}$ με συντελεστή σύζευξης συμβατό με το Καθιερωμένο Πρότυπο (ιστόγραμμα) και σύγκριση αυτής με ξεχωριστούς υπολογισμούς για διακριτές παραμέτρους (κόκκινα σημεία και γραμμή). [Figure 12 στο κυρίως κείμενο.]

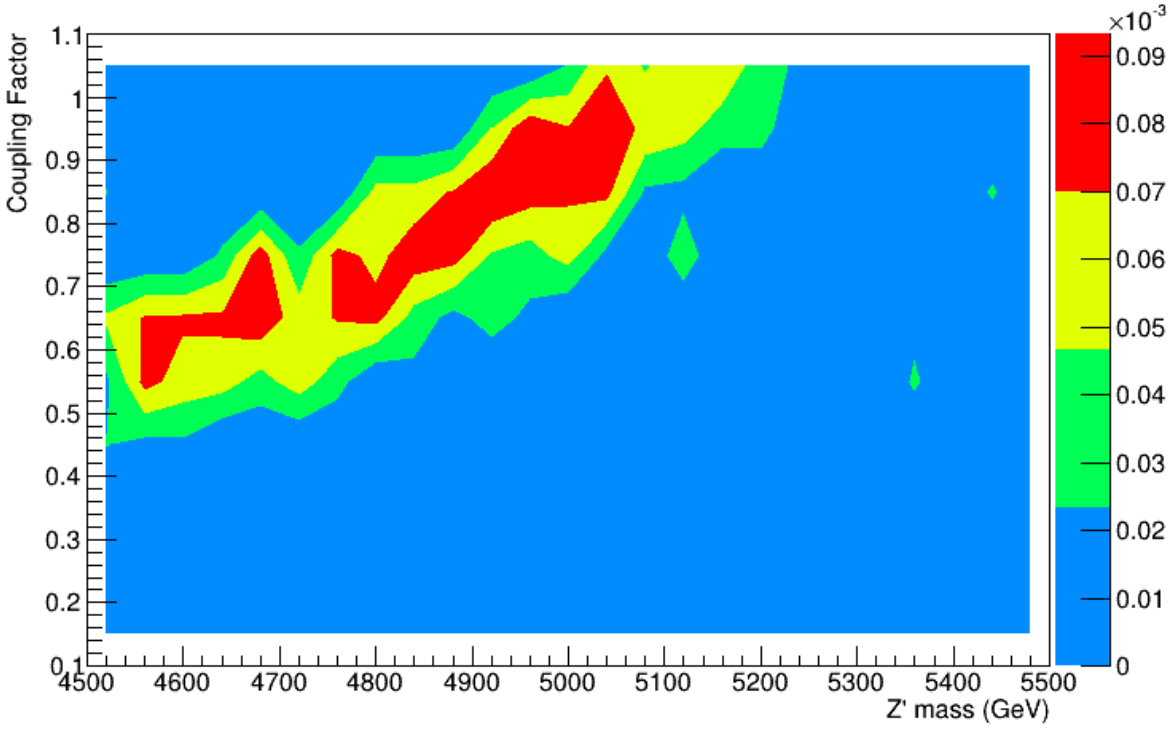
Τέλος, παράγουμε το δισδιάστατο διάγραμμα αποκλεισμού σε συνάρτηση της μάζας του Z΄(εύρος τιμών 3000 GeV με 5500 GeV) και του παράγοντα σύζευξης (εύρος τιμών 0.1 με 10) και συγκρίνουμε με το αντίστοιχο διάγραμμα που παράξαμε με τον παραδοσιακό τρόπο. Η σύγκριση φαίνεται στο Σχήμα ΣΤ. Παρατηρούμε ότι επιτυχαίνουμε συνεπή αποκλεισμό αλλά με μεγαλύτερη διακριτική ικανότητα και χωρίς επανάληψη προσομοίωσης για κάθε συνδυασμό τιμών παραμέτρων. Ο θόρυβος που εμφανίζεται είναι στατιστικός και μπορεί να περιοριστεί με παραγωγή περισσότερων δεδομένων. Επίσης παρατηρούμε ιδιαίτερα καλύτερη διακριτική ικανότητα στον προσδιορισμό της καμπύλης αποκλεισμού (της καμπύλης που διαχωρίζει τις δύο περιοχές) που στο προηγούμενο διάγραμμα θα έπρεπε να βρεθεί με χρήση interpolation μεταξύ των ακριανών σημείων αυτών των περιοχών.



Σχήμα ΣΤ : Δισδιάστατη περιοχή αποκλεισμού με μάζα μεταξύ $M_{Z'}$ μεταξύ 3000 και 5500 και συντελεστή σύζευξης μεταξύ 0.1 και 1.1 (κόκκινη περιοχή) καθώς και σύγκριση με τα σημεία αποκλεισμού (μαύρα σημεία) για ξεχωριστούς υπολογισμούς διακριτών παραμέτρων. [Figure 13 στο κυρίως κείμενο.]

Τέλος, παραθέτουμε μία ακόμη δυνατότητα της μεθόδου αυτής σε περίπτωση ανακάλυψης σήματος νέας φυσικής. Στην περίπτωση αυτή, μπορούμε γρήγορα να υπολογίσουμε τον παραμετρικό χώρο που είναι συνεπής με την ανακάλυψη. Δηλαδή τους συνδυασμούς των ελεύθερων παραμέτρων (της μάζας του Z' και του συντελεστή σύζευξης στην περίπτωση μας) που μας δίνουν την τιμή της ενεργού διατομής και τις κινηματικές τιμές που παρατηρήσαμε στο πείραμα. Στο Σχήμα Ζ παραθέτουμε ένα διάγραμμα όπως και στο Σχήμα ΣΤ όμως αυτή τη φορά

παρουσιάζουμε την περιοχή που αντιστοιχεί σε τιμές της ενεργού διατομής συμβατές με μία υποθετική μέτρηση ανακάλυψης με τιμή ενεργού διατομής 0.022 ± 0.006 fb. Η περιοχή του παραμετρικού χώρους που μας δίνει ενεργές διατομές συμβατές με την παραπάνω μέτρηση, εντός σφάλματος, έχει τη μορφή τόξου.



Σχήμα Z : Συνδυασμοί της μάζας $M_{Z'}$ και συντελεστή σύζευξης συνεπείς με μια υποθετική ανακάλυψη με παρατηρήσιμη ενεργό διατομή 0.022 ± 0.006 fb. [Figure 14 στο κυρίως κείμενο.]

Συμπερασματικά, καταλήγουμε πως πράγματι η μέθοδος αυτή είναι αποτελεσματική καθώς μας προσφέρει τον προσδιορισμό της περιοχής αποκλεισμού θεωριών με ελεύθερες παραμέτρους ή προσδιορισμό των συνδυασμών αυτών των παραμέτρων σε περίπτωση ανακάλυψης, με καλύτερη διακριτική ικανότητα από τις υπάρχουσες διακριτές μεθόδους. Η μέθοδος αυτή μπορεί να αποδειχθεί ιδιαίτερα χρήσιμη σε μελέτες θεωριών με πολλές ελεύθερες παραμέτρους, όπως θεωρίες Υπερσυμμετρίας.

List of Figures

List of Tables

Introduction

A common way of simulating particle-physics events is sampling the kinematic phase space during Monte-Carlo integration with importance sampling. If the integrand is a cross section, the momenta of the outgoing and incoming interacting particles will be generated in phase space regions that are more probable, which leads to realistic simulation. Although this method of particle generation is sufficient in Standard Model processes, it is inefficient in new-physics simulations, where the theories have a number of unknown parameters. In the latter cases, the parameters have to be set to fixed values each time simulated events are produced.

The purpose of this thesis is to demonstrate the feasibility of using an adaptive integrator of cross section to sample not just the kinematic space (particle momenta) but also the parameter space of a theory. This will allow the use of this integrator as a new-physics particle generator in a multidimensional range of theoretical parameters. A single simulation sample will include all variations of parameters, generated in parameter-space that is more probable.

This will allow us to set limits on parameters described by theories in the absence of discovery or, in case of discovery, be able to make quick determinations of the parameter space that agrees with said observation. At the same time, all correlations between particle momenta and theory parameters are included in a single file, which allows easy optimization of kinematic event selection for maximizing new-physics signal.

The thesis is organized as follows:

In Chapter 1 we outline the theoretical background for the physics process we study. Starting from the Standard Model we will introduce the Feynman calculus which we will utilize to obtain the expression for the Drell-Yan process. The new-physics process we study will be Z' boson production, which is the Drell-Yan process with a much larger pole mass and a variety of coupling strengths to leptons. We also briefly explain the statistics of setting limits to theories with a confidence level of 95% in some simplified experimental situations.

In Chapter 2 we outline the computational tools used, most importantly the VEGAS adaptive integration algorithm with importance sampling and the Madgraph matrix element generation.

In Chapter 3 We describe the main program (Apollo) which expands the integration to parameters of the theory. We demonstrate the usage of Apollo by efficiently calculating Z' production cross sections with subsequently decays to electron-positron pairs. By including the Z' mass as an integration parameter, thus a sampled quantity, we reproduce the distribution of cross section as a function of mass. When we add the coupling factor as an integration variable, we can set exclusion limits and determine discovery parameters in the 2D parametric space of mass vs coupling factor.

In Chapter 4 we summarize the conclusions.

Chapter 1: Theory

The Standard Model

High Energy Physics or Particle Physics [1] is a field of physics that describes the basic constituents of the world and their interactions. Although philosophical traces can be found in ancient times, for example Democritus' atomic theory, the beginning of particle physics is usually attributed to J.J. Thomson with his discovery of the electron in 1887, the first truly fundamental particle discovered. This lead to a cascade of new particles being discovered in the 20^{th} century many of them, as was proven along the way, were blocks of even smaller fundamental particles (for example quarks that make up the proton and neutron in the nucleus of every atom). The culmination of all these efforts from many great scientists is what we now call the Standard Model (SM) of elementary particle physics. The Standard Model includes all our experimentally verified knowledge of particle physics. According to the Standard Model all fundamental particles can be categorized into either fermions (half-integer spin) or bosons (integer spin).

The fermions are the constituents of the matter that makes up the world around us. They come into three generations and each generation includes two quarks and two leptons for a sum of twelve fundamental fermion particles. These are shown in Table 1. All of chemistry and most of what we see around us can be attributed to the particles of the first generation. The rest of the particles only come into existence, in at least an observable way, under high energy processes such as high energy physics experiments, supernovas, cosmic ray bombardment into our atmosphere and similar phenomena. But even when they come into existence they are short-lived and only exist for a fraction of a second before decaying into more stable particles.

The bosons are the particles responsible for the interactions between particles giving rise to three forces of nature, the electromagnetic force, the weak nuclear force and the strong nuclear force. Yet we know that there is also a fourth force, gravity. The inclusion of gravity into the Standard Model is a highly researched topic with many theoretical attempts as well the searches for the elusive graviton. But there is no experimental evidence to support these claims till this day. These bosons are also known as force carriers and can be seen in Table 2. The most well-known of them is of

course the photon, which is the carrier of the electromagnetic force, followed by the W^+, W^- and Z^0 bosons, carriers of the weak nuclear force, and eight gluons, which are the carriers of the strong nuclear force. The most recent addition to the Standard Model is the Higgs boson, discovered at the Large Hadron Collider (LHC) in 2012, which is a scalar boson responsible for the mass of the elementary particles.

	Fermions		
	1st Generation	2nd Generation	3rd Generation
Quarks	Up (u) Down (d)	Charm (c) Strange (s)	Top (t) Bottom (b)
Leptons	Electron (e) Electron-Neutrino (ν_e)	Muon (μ) Muon-Neutrino (ν_μ)	Tau (τ) Tau-Neutrino (ν_τ)

Table 1 : The Fermions of the Standard Model.

	Bosons	Interaction Mediators	Relative Magnitude of Interactions*
Gauge Bosons	Photon (γ)	Electromagnetic Interaction	10^{-2}
	Gluons (g)	Strong Nuclear Interaction	1
	Z-Boson (Z^0) W Bosons ($W^{\pm}$)	Weak Nuclear Interaction	10^{-7}
Scalar Boson	Higgs Boson (H)	Higgs Field	-

Table 2 : The Bosons of the Standard Model.

*This is the relative magnitude between two protons when they are just in contact. There is also the gravitational interaction with a relative magnitude of 10^{-39} but no mediator (graviton?) has yet been experimentally detected or verified.

Finally to complete the listing of all the fundamental particles, we shall add the antiparticles, which is another class of particles discovered from the late 1930s to late 1950s. These particles have the same attributes as the "normal" particles above but some quantum properties reversed such as charge and also when they come into contact with their counterpart will annihilate each other. The prime example is the positron which is the anti-electron, the antiparticle counterpart to the electron. This holds true for nearly all particles except for the photon, gluons, the Z boson and the Higgs boson which are their own antiparticle. So if we add them all up we have 30 fundamental particles, 24 fermion particles including their antiparticle counterparts as well as 6 bosons which are their own antiparticles. Still the search goes on.

From a mathematical perspective [2-4] the Standard Model is a gauge theory that utilizes the Lagrangian formalism in order to describe the fields (that correspond to the three forces of nature) and their interactions. The Lagrangians are dictated to by the symmetries that each force adheres to.

The gauge theory of the Standard Model is the product of 3 gauge symmetries

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$$

$SU(3)_C$ for the strong interaction

$SU(2)_L$ for the weak interaction

$U(1)_Y$ for the electromagnetic interaction

The corresponding compacted Standard-Model Lagrangian can be written as:

$$\mathcal{L}_{SM} = -\frac{1}{4}\boldsymbol{B}_{\mu\nu}\boldsymbol{B}^{\mu\nu} - \frac{1}{4}\boldsymbol{W}_{\mu\nu}\boldsymbol{W}^{\mu\nu} - \frac{1}{4}\boldsymbol{G}_{\mu\nu}{}^{i}\boldsymbol{G}^{\mu\nu i} + \sum_i \bar{f} i\gamma^\mu D_\mu f \quad \textbf{(Eq. 1)}$$

Where "$\sum_i \bar{f} i\gamma^\mu D_\mu f$" is sum over all fermions and D_μ is the overall covariant derivative:

$$D_\mu = \partial_\mu + ig_1\frac{Y}{2}B_\mu + ig_2\frac{\tau_i}{2}W_\mu^i + ig_3\frac{\lambda_i}{2}G_\mu^i$$

First Term: $B_{\mu\nu}$ is the gauge field from the electromagnetic $U(1)_Y$ symmetry.

Second Term: $W_{\mu\nu}$ is the gauge field derived from the $SU(2)_L$ symmetry of the weak interaction.

Third Term: $G_{\mu\nu}$ is the gauge field derived from the $SU(3)_C$ symmetry that describes the stong interaction.

Fourth Term: D_μ is the interactions between fermions through fields

The Lagrangian formalism is based on the least action principle. The action for a Lagrangian density $\mathcal{L}$, which is a function of the fields φ_i and its derivatives, is given by:

$$S = \int_{t_i}^{t_f} \mathcal{L}(\varphi_i, \partial\varphi_{i\mu}) \quad \textbf{(Eq. 2)}$$

The least action principle requires that under small variations the action above must not change:

$$\delta S = 0$$

And so

$$\delta \int_{t_i}^{t_f} \mathcal{L}(\varphi_i, \partial\varphi_{i\mu}) = 0,$$

which gives the Euler-Lagrange equation:

$$\frac{d}{dt}\left(\frac{\partial\mathcal{L}}{\partial(\partial\varphi_{i\mu})}\right) = \frac{\partial\mathcal{L}}{\partial\varphi_\iota} \qquad \textbf{(Eq. 3)}$$

This gives us the equation of motion of the physical system under study.

This formalism follows the classical mechanics formalism but in quantum field theory the Lagrangians are regarded as axiomatic functions whose only constrain is that they produce the field equations that are consistent with experimental observations. They can be multiplied by a scale factor or have an arbitrary factor added to them but when entered the Euler-Lagrange equations must result in the same field equations.

Feynman Calculus

Feynman diagrams are a convenient way to represent graphically the interactions between particles and fields. In this formalism, straight lines represent fermions while wavy, curly or broken lines represent bosons. The arrows indicate the time direction with time flowing from left to right (there is also the down-up approach used in much bibliography but we shall adhere to the left-right flow of time for the remainder of this project). If in some diagram a particle is depicted as travelling backwards in time, it is the equivalent of the corresponding antiparticle moving forward in time, and vice versa. The particles that correspond to lines which enter or leave the diagram are the "real" particles that we observe whereas the particles that join these particles in between are called virtual particles and only exist for the duration of the interaction.

Fermions and bosons meet at vertices where charge, energy, and momentum are conserved and the strength of the interaction is represented by a coupling constant (for example for electrodynamic interactions the coupling is $a = \frac{e^2}{4\pi\hbar c} \cong \frac{1}{137.0360\ldots.}$).

These diagrams are very useful as we can easily calculate the amplitude *M* of a given interaction by using Feynman's rules and eventually arrive at the cross section and decay rate of the interaction by using the Golden Rule of Fermi for scattering:

$$\frac{d\sigma}{d\Omega} = \frac{1}{4\pi^2}|M_{fi}|^2 \left[\frac{E_a E_b E_c E_d}{(\hbar c)^4 s}\right]\left(\frac{p_f}{p_i}\right) g_f \qquad \textbf{(Eq .4)}$$

This gives us the cross section for a process a+b $\rightarrow$ c+d in the centre of mass frame. The total center of mass energy is given by $\sqrt{s} = E_a + E_b = E_c + E_d$. $|M_{fi}|$ is the matrix element that is found using the Feynman rules, p_f and p_i are the momenta of the final and initial state respectively. E_x is the energy of each particle and g_f is the statistical weight of the final state spins. Our project utilises these rules together with Fermi's Golden Rule to arrive at the expression of the Drell-Yan process's cross section, which we will compute.

The most basic diagram for a particular process is called a Leading Order diagram for that process and each subsequent diagram that involves more vertices are call Next to Leading (NLO) order and so on. Every next order that includes more vertices is more and more unlikely. This can be understood for example from the coupling constant mentioned above, as each additional vertex adds another 1/137 factor making the procedure less likely by this amount. In our project we will only use the LO diagram of the Drell-Yan process.

Experimental Measurement of Cross Sections

Cross sections together with decay rates are two quantities that are regularly measured in experiments. Cross sections are fairly popular among scattering experiments such as the LHC [6], where we have the collision of two proton beams with well-defined momenta. The likelihood of any particular final state can be expressed in terms of a cross section. Roughly speaking it is the effective area of a chunk taken out of one beam, by each particle in the other beam, which subsequently becomes the final state we want to observe. Each cross section is also unique to each given process.

This can easily be observed in a collider experiment such as the CMS [5] at the LHC, where we know that the number of events (*N*) observed is

$$N = \sigma \times L \times \varepsilon \times a, \qquad \textbf{(Eq. 5)}$$

where σ is the cross section we want to measure, L is the integrated luminosity, a measure of how many collisions took place during the duration of data taking. This is different for every experiment and depends upon the accelerating apparatus. At the LHC the instantaneous luminosity is typically about 10^{34} $cm^{-2}s^{-1}$. The integrated luminosity of Run-II is 139 fb^{-1}, which is the value we use in this thesis. The efficiency ε and acceptance a correspond to the efficiency and acceptance of the detector that we use. In the scope of our work we will assume that our detector is a perfect detector that can detect every single event generated so both the acceptance and efficiency will be hitherto regarded as equal to one ($\varepsilon \times a = 1$). From the above we conclude that the cross section at a perfect detector can be measured as the number of scattering events divided by the integrated luminosity.

The Drell-Yan Process

Now we shall compute the cross section using the Feynman formalism for the Drell-Yan process [4, 7] which is a process in which a quark and antiquark pair annihilate each other and produce a lepton pair which are the final observable state.

$q\,\bar{q} \rightarrow l^{-}l^{+}$, as shown in Fig. 1.

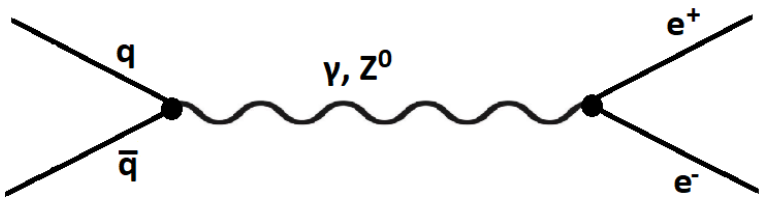


Figure 1 : Leading-Order Feynman diagram for the Drell-Yan progress. Note we have already only taken into account the electrons and not all leptons.

This process was first proposed by Sidney Drell and Tung-Mow Yan in 1970 [8] to describe the production of lepton-antilepton pairs in high-energy hadron collisions. Since then it has been experimentally and theoretically scrutinized as it proved a very useful tool in further understanding QED but also QCD. The Drell-Yan process was initially proposed as a process where two quarks annihilate into a photon which later produces the pair of leptons. With later advancements and the understanding of electroweak interactions we now know that the intermediate boson can be either a photon or a neutral boson Z^0. One important aspect is that it can by described as a factorized hadronic cross section for the massive dilepton production in hadronic collisions, which can be expressed as a sum (over quark pairs) of products of partonic hard parts and universal PDFs which depend on the properties of the colliding hadrons. The Drell-Yan process also proved valuable for the discovery of of the W and Z by bosons by the UA1 and UA2 experiments as well as aiding in the discovery of the three heavy quarks c, b and t. Other important contributions are the probing of the antiquark contents in hadrons, as well as the parton structures of pion, kaon and antiproton. One main feature found and measured by the Drell-Yan process (while comparing to similar DIS measurements) is also the sea antiquark distributions in heavy nuclei to which the Drell-Yan process is particularly sensitive.

Something important to note is that our process will be conducted within the confines of the LHC experiment, which means that our initial quark and antiquark will be part of composite particles namely the protons. The protons are baryons composed of three quarks, two up quarks and one down quark. They have a +|e| charge and mass close to one atomic mass. They are the most stable hadrons and together with the neutrons and electrons make up the vast majority of matter in the universe. It is also the nucleus of hydrogen so it is very easy to isolate. But apart from the three quarks which are called valence quarks, a proton also contains gluons that are responsible for keeping the quarks within the proton. Because of the nature of QCD, the proton also contains an indefinite number of virtual "sea" quarks. Unfortunately, the structure of a proton as we see is quite complicated and so we still do not have a complete theoretical picture of how the momenta of its components are distributed. Luckily, we have a fairly accurate experimental description, in the form of PDFs (Particle Distribution Functions). These functions are determined from fits of observed quantities from many different experiments, both collider and deep inelastic scattering experiments. The structure of the proton is important when calculating the cross section, because an antiquark coming from a proton can only be a sea antiquark. Also we have to take into account the contributions from the spectator quarks (the quarks that don't take immediate part in the interaction but by their presence alter the final cross section). We will also examine only the electron-positron pair production from the possible lepton pairs for simplicity. So we can rewrite the Drell-Yan process according to our constraints as:

Examined Process: $p\,p \rightarrow e^- e^+$

which corresponds to the Feynman diagrams (in LO) shown in Fig. 2.

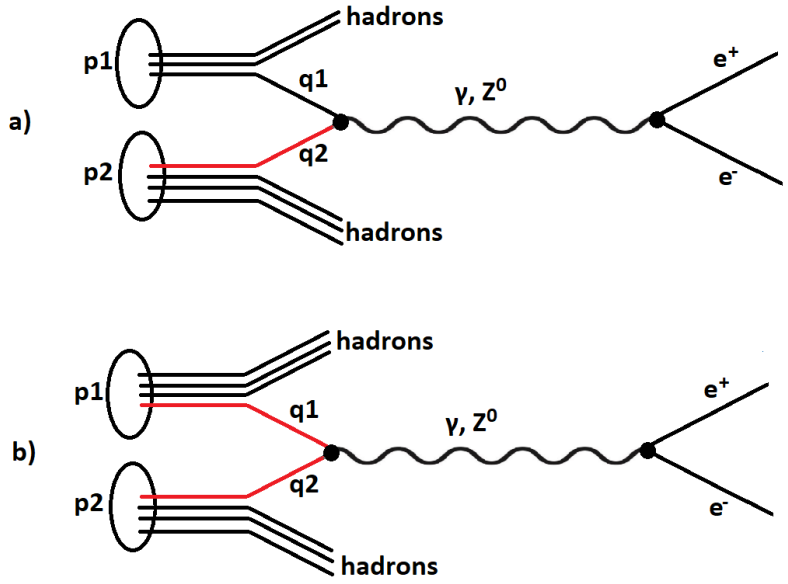


Figure 2 : Leading Order Feynman Diagrams for the Drell-Yan interaction we will be investigating. Red line denotes a quark coming from sea.

In the above Feynman diagrams we have noted with p1 and p2 the two colliding protons, q1 and q2 their corresponding quarks that take immediate part in the Drell-Yan process and with "hadrons" the corresponding byproducts of the protons after the collision. We have also noted in red color the quarks that originate among the sea quarks. Note that the only difference between diagram a) and

diagram b) is the origin of the q1 quark: in diagram a) it is one of the valance quarks of the p1 proton whereas in diagram b) it originates from the sea quarks of the p1 proton just like the q2 quark from the p2 proton. This does not mean that they are different processes, as in both cases we have two protons that produce a lepton pair, but has it has to be taken into account when determining the matrix elements. This second diagram also takes into account the scenario of having the antiquark q1 interact with the quark q2. Also it should be added that q1 and q2 include all the different pairs of quarks-antiquarks and vice versa including up, down, charm, strange and bottom quarks with their corresponding antiquarks (the top quark is omitted as its mass is very large and so it cannot be found inside the proton).

For the purposes of our project we will restrict ourselves to Leading Order processes but will compensate for this later on with a correction on the final cross section computed.

Now by simply applying Feynman's rules for QED and utilizing Fermi's Golden Rule for cross sections we can arrive at an expression that gives us the cross section of the above examined Drell-Yan process:

$$\sigma = \int \sum_{q_i,q_j} f_i(x_i,\mu) f_j(x_j,\mu) \left|M(P_i,P_j,P_{out,k})\right|^2 \delta^4(P_i + P_j - \sum_k P_{out,k}) \prod_k \frac{dP_{out,k}}{E_k} \frac{dx_i dx_j}{E_i E_j} \quad \textbf{(Eq. 5)}$$

Which is the expression that use in our software to find the cross section for the interaction. f_i (f_j) is the PDF function which gives us the probability that the the quark i (j) has fraction x_i (x_j) of the proton's momentum, for the energy scale μ, $P_{i,j}$ are the momenta of colliding quarks, $P_{out,k}$ the momenta of outgoing lepton and $E_{i,j,k}$ the corresponding energies. M is the matrix element, for which we shall utilize Madgraph to get an expression for. The delta function guarantees the conservation of momentum. We also show the integration phase-space factors that we also compute. In equation (5) the matrix element squared multiplied by the delta function and phase-space energy factors gives us the partonic cross section, whereas its multiplication with the PDFs and the sum over all partons gives us the hadronic integrand. The integration of the hadronic integrand over all momenta gives us the total hadronic cross section.

Limits on New Theories

The Standard Model of Particle physics is up to today our most successful theory, able to explain a large number of phenomena. Yet the theory has certain shortcomings, things that up to today it cannot explain. Some of these are the dark matter, dark energy, the neutrino's mass, matter-antimatter asymmetry, strong CP violation, the hierarchy problem and more. So this means that there is plenty of room for new physics and theories beyond the Standard Model.

When experimentally testing a theory, in the absence of observation, limits on the theoretical parameters are set. For example, if a theory predicts the existence of a particle and no signal is experimentally observed, we can make a statement about what is the highest cross section of the process that could generate this particle, which usually translates to the lowest mass the particle can have if it exists.

It can be statistically shown from a Poisson distribution that if we expect zero background, we perform the experiment and we count zero events (an event could be the creation of said particle),

then we can state with a 95% confidence level (CL) that the theory cannot produce three (3) or more events.

Poisson distribution: $P(X \leq x) = \sum_{r=0}^{x} \lambda^r \frac{e^{-\lambda}}{r!}$ **(Eq. 6)**,

where λ is the mean and r is the number of events. If we substitute the values r=0 and λ=3 into the Poisson expression (which corresponds to the probability of finding zero events in a distribution with mean 3) we will find that it is equal to:

$$P(\lambda = 3, r = 0) = 0.0498 \approx 5\%$$

This can be represented graphically in Fig. 3.

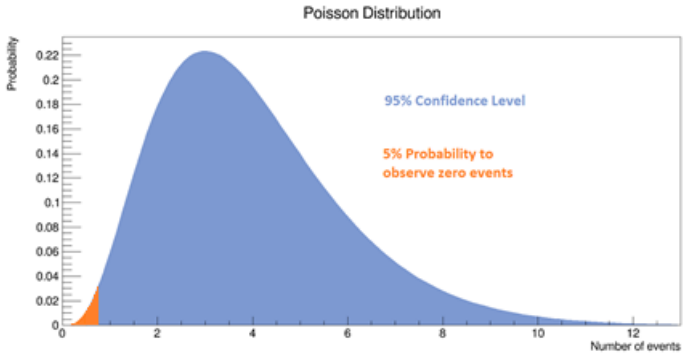


Figure 3 : Poisson distribution (in continuous-function approximation) showing the 5% chance of finding zero events.

The 95% CL limit on the expected signal events can be translated to a 95% CL cross section upper limit if divided by the integrated luminosity L. If no signal events are observed and no background events are expected, we exclude cross sections above 3/L, at 95% CL.

Chapter 2: Computational Tools

VEGAS

The VEGAS algorithm [9], invented by Peter Lepage in 1978, is a multidimensional integration algorithm. It is an iterative and adaptive Monte Carlo scheme and is considerably more efficient than other algorithms especially in high dimensions ($n \geq 4$). It is widely used for multidimensional integrals that occur in elementary particle physics. It utilizes both importance and stratified sampling techniques and is ideal for computing the scattering amplitudes derived from Feynman perturbation theory.

The main benefits of the integration as stated by LePage [9] are

(a) A reliable error estimate for the integral is readily computed.
(b) The integrand doesn't need to be continuous.
(c) The convergence rate is independent of the dimension of the integral
(d) It is adaptive.

We should also add that

(e) It avoids the "K^d explosion". The overhead and storage requirements grow only linearly with dimension, because during sampling we end up with separable distribution densities [10].

Characteristics (a), (b) and (c) are common among Monte Carlo methods but what VEGAS adds is adaptation (d) which sets it apart from the others. One of the major problems in multidimensional integration is the exponential growth with the increasing dimension of the integration volume over

which the integrand is computed. But VEGAS solves this by using sampling techniques before the actual integration. These sampling techniques prepare the integrand so it can be more efficiently integrated by the Monte Carlo that will follow.

So, there are two steps to the VEGAS algorithm. The first step includes two sampling techniques, importance sampling and stratified sampling. These apply an automatic transformation to the integration variables in order to flatten the integrand which makes the integrand easier and improves the estimate. Then, in a second step VEGAS, computes the transformed integrand and produces a Monte Carlo estimate.

Many variants of VEGAS exist today. The one used by us is supplied in Numerical Recipes [10] and is written with FORTRAN 77 (see Appendix B).

The inputs that VEGAS requires are (using the variable names given):

region: multidimensional rectangular volume in which the integrand is to be computed,

ndim: number of dimensions,

itmx: number of statistically independent evaluations of the desired integral,

(It should be noted that although statistically independent the previous iterations assist the next iterations by refining the sampling grid as will be explained later on.)

ncall: the number of function evaluations per each integral evaluation (*itmx*),

init: input flag which signals whether the call is a new start or a subsequent call for additional iterations,

fxn: user-supplied integrand function,

nprn: input flag that controls the amount of diagnostic output, usually =0.

And its outputs are:

tgral: the final estimate of the integral,

sd: standard deviation of the above integral,

chi2a: Quantity that evaluates the quality of the results, if this is significantly larger than 1 then this means that the results of the iterations are statistically inconsistent and the results are suspect.

VEGAS employs two sampling strategies in order to prepare the integrand before the actual integration. These are the “importance sampling” and the “stratified sampling”.

Importance Sampling:

Importance sampling is the principle adaptive strategy employed by VEGAS. This technique focuses the integration around the regions where the integrand has its largest values (for example around peaks). It accomplishes this by choosing transformations for each integration variable that minimize

the statistical errors in Monte Carlo estimates whose integrand samples are uniformly distributed in the new variables. As explained in [11], the idea in one-dimension replaces the original integral over *x*

$$I = \int_a^b dx f(x) \quad \textbf{(Eq. 7)}$$

With an equivalent integral over a new variable *y.*

$$I = \int_0^1 dy\, J(y) f(x(y))$$

Where *J*(*y*) is the Jacobian of the transformation. As a result, a simple Monte Carlo estimate of the transformed integral would be given by

$$I \approx S^{(1)} = \frac{1}{M} \sum_y J(y) f(x(y))$$

Where the sum is over *M* random points uniformly distributed between 0 and 1. This estimate $S^{(1)}$ is also a random number from a distribution whose mean is the exact integral with a variance of

$$\sigma_I^2 = \frac{1}{M}\left(\int_0^1 dy\, J^2(y) f^2\big(x(y)\big) - I^2 \right) =$$

$$\frac{1}{M}\left(\int_a^b dx\, J^2(y(x)) f^2(x) - I^2 \right)$$

The standard deviation σ_I^2 is an estimate of the possible error in the Monte Carlo estimate and is constrained by

$$\int_a^b \frac{dx}{J(y(x))} = \int_0^1 dy = 1$$

And through some calculations is shown to be minimized if

$$J\big(y(x)\big) = \frac{\int_a^b dx\, |f(x)|}{|f(x)|}$$

It can be easily shown that this kind of transformations minimize the standard deviation in areas where the integrand has high peak values, since

$$\frac{1}{J} = \frac{dy}{dx} \propto |f(x)| \,.$$

The end result is that regions in *x* space where |*f*(x)| is large are stretched out in *y* space. It is the equivalent of changing the size of "bins" during the integration in order to have more samples around the peaks than in areas with smaller (or zero) values of integrand. This means that when we later apply the uniform Monte Carlo integration on the transformed *y* space we will have many more measurements concentrated around regions where we have peaks than if we were to integrate in

the pre-transformed *x*-space. It should be noted that the distribution of the Monte Carlo estimates becomes Gaussian for a very large number of *M*.

The result of the above process in the VEGAS grid. As we typically have no knowledge of the integrand initially, we start with a uniform *x* grid. The program makes estimates for the integral while it samples the integrand refining its grid with every iteration, which affects the next one. After several iterations, the number of which depends on the integrand complexity and the number of dimensions, the VEGAS grid converges to its optimal configuration.

The algorithm can be generalized to any number of dimensions by applying the same procedure in every dimension, making grid increments along an axis smaller in regions where the projection of the integral on the axis has a larger value and larger where it has lower values.

For example, Fig. 4 shows the grid that was computed for the evaluation of a two-dimensional Gaussian integral.

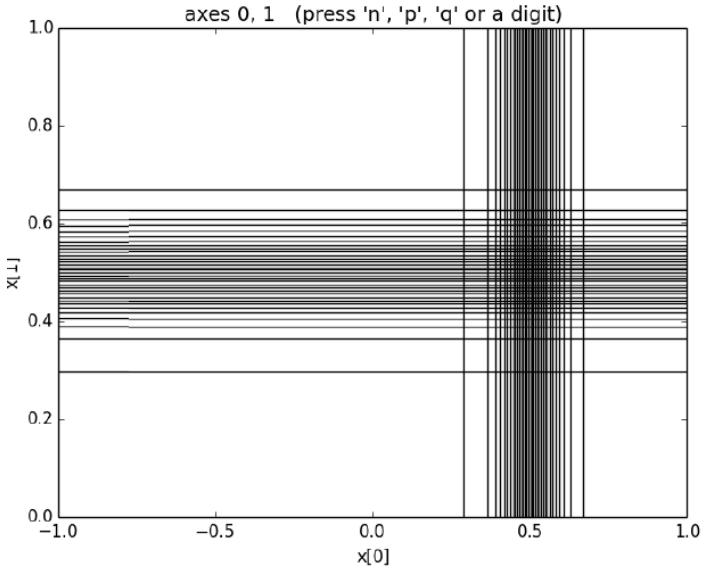


Figure 4 : Optimal grid for the integration of a two-dimensional Gaussian produced with Importance Sampling. As shown in [11].

Stratified Sampling:

The above importance-sampling technique has its shortcomings. One prime example are the so-called phantom peaks that may be created. This occurs when we generalize into higher dimensions because we examine the dimensions separately. This can create overlapping regions where we assign samples as though we had a peak in this region when in reality they just share the coordinates of different but unrelated peak regions. For example the x coordinate of one peak and the y coordinate from a second peak. This is most prevalent when the peak distribution happens to be along diagonals of the integration volume. We demonstrate this using the example in [11] as shown in Fig. 5.

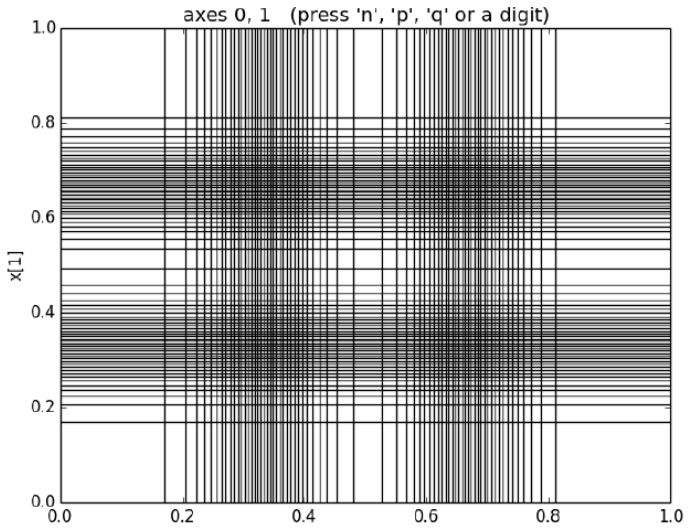


Figure 5 : Optimal grid for the integration of two 2-dimensional Gaussian produced with importance sampling showing phantom peaks. As shown in [11].

The grid of Fig. 5 has been created using importance sampling for two four-dimensional Gaussian integrals whose means lie at [0.33, 0.33] and [0.67, 0.67] respectively. The above grid shows that VEGAS concentrates its sampling in the regions around [0.33, 0.33] and [0.67, 0.67], where the peaks are, but unfortunately, also concentrates in regions around [0.67, 0.33] and [0.33, 0.67], where the integrand approaches zero.

Although this grid still produces better results than a uniform grid, it is apparent that useful resources are being wasted as computational power is directed away from our regions of interest into these phantom peaks. This is why in later alterations of the VEGAS algorithm, such as the one we utilize, stratified sampling is used as well as importance sampling. Stratified sampling focuses the sample region not according to its largest value but based on its variance or standard deviation in our case. This is generally useful when we want to focus our computations in regions where we have a greater uncertainty, in order to get a clearer result, instead of regions where we have a lower uncertainty and fewer computations are needed to arrive to a satisfactory result.

Consequently, VEGAS divides the d-dimensional y-space volume into M_{st}^{d} hypercubes using a uniform y-space grid with M_{st} stratifications on each axis. It estimates the integral by doing a separate Monte Carlo integration in each of the hypercubes, and with adding the results together provides an estimate for the integral over the entire integration region. But here we adjust the number of evaluations used in a hypercube in proportion to the standard deviation in one iteration in order to set the number of evaluations for that hypercube in the next iteration. The net effect is that we focus our computations in areas where the potential statistical errors are largest.

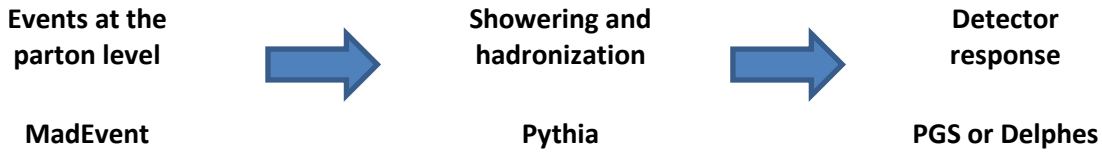
This overall directs our computational efforts away from the above-mentioned phantom peaks and focuses them in the regions where the real peaks lie as these real peaks are the regions producing most of the error contributions of the integrand.

If the integrand of VEGAS is a matrix element, as in our case, then the importance sampling of particle momenta can lead to a particle generator. Because we preferentially select momenta that maximize the partonic cross-section, which are the most probable combinations of momenta. To

transform VEGAS to a particle generator, we have to store the momenta it samples during integration of the partonic cross section after the creation of the grid.

Madgraph

MadGraph [12] is a Monte Carlo event generator used in collider physics to simulate events detected by many experiments such as the LHC. Madgraph integrates partonic cross sections (based on matrix elements it generates) using VEGAS. MadGraph software can be also coupled to a number of programs which allow for a complete simulation at the LHC, going from events at the parton level to detector signals. It is available online and part of the MadGraph5_aMC@NLO project. We should note that in realistic settings we could use MadGraph to examine further complex processes that follow the creation of particles within the CMS detector such as hadronization and showering or detector inefficiencies, as shown below.



For the scope of our work, however, we do not need the Pythia showering and hadronization because we are investigating only leptonic final states. Furthermore, we do not utilize a detector simulator because we are investigating a perfect detector. For the needs of our project we only use MadGraph for the generation of the Drell-Yan process matrix elements at leading order. Furthermore, we do not need a detector simulator: our goal is to demonstrate that the method works; the detector acceptance and efficiency do not affect our conclusions.

MadGraph provides us the matrix elements for the Drell Yan process $pp \rightarrow e^+e^-$, for all quark combinations.

Parton Distribution Functions and Helicities

In order to get the hadronic integrand to be integrated by VEGAS, all quark combinations have to be summed for all allowed spins. For that reason we need to access the parton distribution function of the proton and the helicity combinations. We do that through CERNLIB and DHELAS libraries. We use the Cteq5 PDF, as implemented in CERNLIB [13]. The function *Ctq5pdf(Iparton,X,Q)* returns the parton distribution inside the proton for parton *Iparton* at Bjorken_X equal to *X* and scale *Q* in the PDF set. We use an older PDF but this does not affect the scope of the thesis.

ROOT

Root [14] is a framework for data processing created at CERN that is widely used today by physicist especially in the field of high energy physics. It is written mainly in C++ and is used for the analysis of large amounts of multidimensional data. For the purposes of this project it has been installed and used within a Linux environment.

Root allows us to save, access, mine and publish data as well as run our own applications and also provides a data structure, the tree, which allows fast access of huge amounts of data many times faster than they would have been accessed otherwise.

The data analysis, plots and histograms in this thesis have been done using ROOT.

Chapter 3: Apollo new-physics generator

Drell-Yan Implementation

Before we investigate the expansion of the code to new physics, we confirm that we can properly simulate Z-boson events as produced through the Drell-Yan process at the LHC. For that reason, we construct partonic cross sections to be integrated by VEGAS, according to equation (5). The matrix elements come from Madgraph and we sum for all quark combinations. The input to the matrix elements are momenta of incoming quarks and outgoing electron-positron pairs, parameterized as shown below.
The incoming quarks have fractions of the incoming protons equal to

$$x_1 = \sqrt{\tau}e^{-y} \quad \text{and} \quad x_2 = \sqrt{\tau}e^{y} ,$$

where τ and y will be used in sampling instead of x_1 and $x2$. We set the center-of mass energy of the colliding protons to be $\sqrt{s} = 13\ \text{TeV}$. This means that the 3-momenta of the incoming quarks in the lab frame are

$$(0,0,x_1E) \quad \text{and} \quad (0,0,-x_2E)$$

where $E = \sqrt{s}/2$

The outgoing particles (electron and positron) have momenta in the center-of-mass frame:

$(Psin\theta cos\varphi, Psin\theta sin\varphi, Pcos\theta)$ and

$(-Psin\theta cos\varphi, -Psin\theta sin\varphi, -Pcos\theta)$,

where $P = \sqrt{x_1 x_2 E^2}$

The leptonic momenta are boosted back to the lab frame.

The incoming and outgoing momenta in the lab frame are arguments to matrix-element subroutines that are generated by Madgraph. The resulted matrix element is multiplied by the probability that the incoming quarks have fractions of the proton's momentum equal to x1 and x2. These probabilities are the parton distribution functions of the proton.

The most efficient choice of kinematic variables integration with VEGAS are:

τ, y, θ, ϕ

with range:

τ: 0 – 1
y: -10 – 10
θ: 0 – π
ϕ: 0 – 2π

From the above it is obvious that the ordinary Drell-Yan process can be generated with a 4-dimensional integration. The main program sets the dimension of the integration and the specifics of the integrand and calls VEGAS adaptive integrator with a large number of iterations (O(1000)) for training of the grid and an equally large number of iterations for generating events. After the grid is determined, the generation of events follows the specific combination of kinematics and parameters that maximize the integrand, i.e., the partonic cross section times the parton distribution functions, summed for all quark combinations, summed/averaged over allowed spins, within constraints of total 4-momentum conservation.

We confirm that we get a reasonable Z-boson production cross section, after we multiply with a K-factor of 1.3 (which corrects for higher-than-leading-order effects), compared to measurements at the LHC. We also confirm that the resulted cross section is very close to the one calculated by Madgraph (~1420 pb with 13 TeV collisions). For the study of the Z boson (or the heavier Z' bosons below) we make sure that we remove the photonic part of the Drell-Yan with a dilepton mass cut.

Apollo

New-physics theories are usually characterized by a number of free unknown parameters. In some theories, such as Supersymmetry, the number of parameters can be fairly large. Ordinary new-signal Monte-Carlo event generation requires fixing all parameters of a theory before generation. Consequently, a large range of parameters and their values is investigated, usually for exclusion limits purposes. A new Monte-Carlo sample has to be generated for each combination of parameter values. For example, if we had 10 parameters and we want to investigate 10 different values per parameter, we would have to generate 10^{10} Monte Carlo samples, which is practically impossible.

For this reason, in theories with a large number of parameters, most of the parameters have to be set to arbitrary fixed values, when limits are set on the free parameters. The Apollo method [15] allows us to set limits for all theory parameters without any artificial fixing of parameter values. This can be done in principle with a single Apollo sample. This is achieved by including the parameters of

the theory as integration variables, when we integrate the partonic cross section. As a result, we sample not only kinematic variables but also parameters of the theory.

Used as particle generator, Apollo produces events that include different kinematics and theory parameters per event. This allows for direct correlations between kinematics and theory parameters and the investigation of a wide range of parameter combinations with a single Monte-Carlo sample. Applications of Apollo include the efficient setting of multidimensional limits on theoretical parameters in the absence of discovery, the optimization of kinematic cuts to study particular regions in parameter space, and quick determination of parameter space consistent with observation in case of discovery. The code structure of Apollo is presented in Appendix A.

We demonstrate the feasibility of Apollo using the Z' boson generation.

Z' Boson Generation

The demonstration of the Apollo concept in this thesis is achieve with a simple example of Z' boson production in 13 TeV proton-proton at the LHC, with subsequent decay to electron and positron pairs. In this thesis, we consider a Z' that differs from the SM Z boson in its mass and coupling strength to lepton. We consider two theoretical parameters that can be altered simultaneously: the mass of Z' and the decay coupling factor, which multiplies both the left-handed and right-handed couplings of Z' to the electrons. For simplicity, and without loss of generality, the following ideal conditions are considered:

a) The detector is perfect (acceptance and efficiency are equal to one),

b) There is no Standard-Model background expectation,

c) There is no observation of any events when we "open the search box".

Usually, the theoretical predictions for the cross section of a new-physics signal come from analytic calculations; the generation of new-physics Monte Carlo is used for the determination of the signal acceptance and efficiency for different new-physics parameters. This acceptance and efficiency affect the cross-section exclusion limit which is translated to limits on the theory parameters. Because of the ideal conditions listed above, as already discussed, the 95% confidence level (CL) exclusion limit is fixed at $3/L$ (where L is the integrated luminosity, 139 fb^{-1} for the Run II of ATLAS or CMS). For that reason, we demonstrate the power of Apollo by using it for the determination of the new-physics cross section, as a function of the new theory parameters. Any combination of parameters is excluded or not, depending on the value of the cross section for these parameters. If the cross section is above the 95% CL limit, the parameter combination is excluded, otherwise it is not.

From the above, it is clear that the demonstration of feasibility greatly depends on the accuracy of the produced cross section, within variations of the parameters. This accuracy is determined by comparing the Apollo-produced cross sections with dedicated runs for fixed theory parameters. In this thesis we do this both in one parameter dimension (cross section as a function of Z' mass) and in two parameter dimensions (cross section as a function of Z' mass and Z' decay coupling factor). Obviously, there may be regions of parameter space with limited Monte-Carlo statistics that lead to

a larger uncertainty in the Apollo calculations. These regions can be identified and a separate run within a different parameter range can be issued. In any case, it makes sense to study the parameter space regions close to expected exclusion limits.

The above consistency checks will demonstrate that the phase-space normalization that should be included in the adaptive integration is well understood. And that the framework can be used to set multidimensional parameter limits. Our goal is to demonstrate that these limits are consistent with the ones that we get from dedicated fixed-parameter runs. But at the same time offer a better resolution in parameter space. Finally, we will demonstrate how Apollo can be used for determining the allowed parameter-space in case of discovery of new physics.

When we move to Z' generation, we add two more integration variables, which are theory free parameters (mass and coupling factor). This requires expanding the integration space to 6 variables and passing the new Z' masses and coupling factors to the matrix elements. The width of the Z' is considered to be the same as that of the SM Z boson. We utilize a dilepton mass cut to remove any photonic part from the dilepton spectrum. To achieve the Z' functionlality we have to modify the ordinary Drell-Yan Madgraph matrix element to include the theory parameters (an example of such modification is shown in Appendix C). We alter the code in 4 steps.

i) First we add the functionality for running Apollo with fixed values of Z' mass (with no alteration of the couplings). Our goal is to reproduce a reasonable SM-like Z' boson cross section as a function of its mass, as a way of validating the cross section produced by Apollo. Fig. 6 shows the published ATLAS Z' limit, which includes a theoretical prediction as a function of mass. Our Z' boson corresponds to the Z'_{SSM} in this plot, if we set the coupling factor to unity.

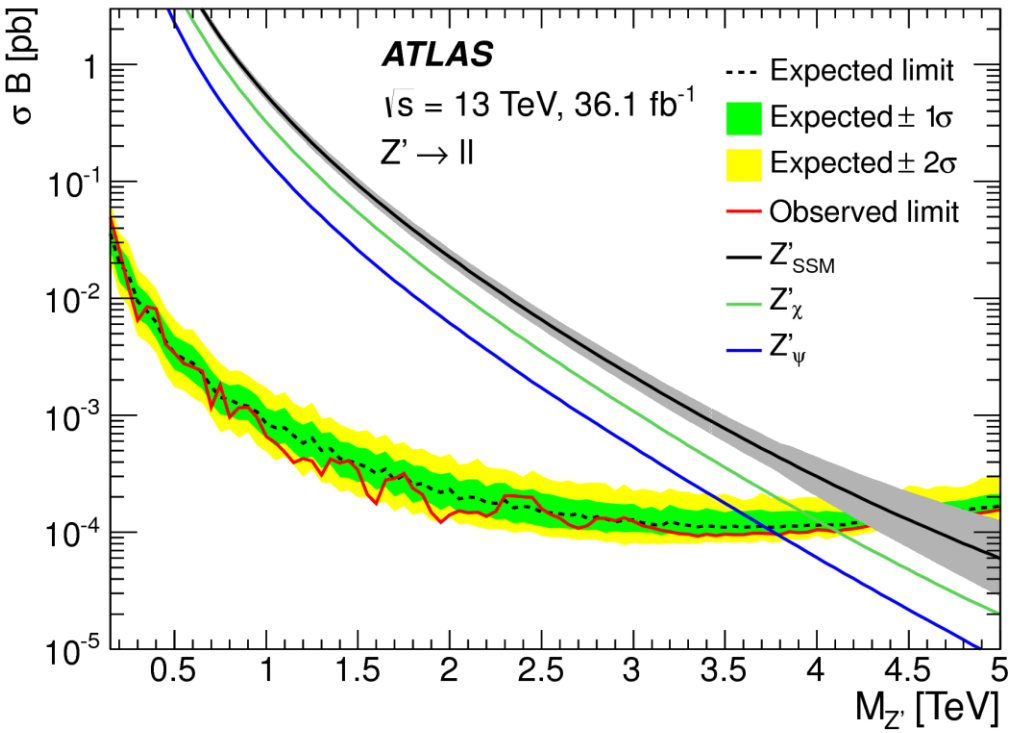


Figure 6 : Published Z' limit by ATLAS [16].

In Fig. 7 we see the comparison of the theoretical cross section as a function of mass (black points, taken from Fig. 6) with the red points which are the results of the Apollo run with a fixed Z' boson mass at a time. Note that there is no integration over this mass yet.

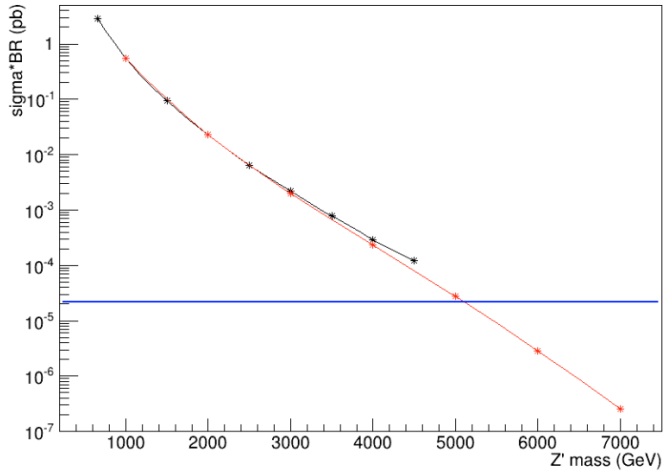


Figure 7 : Comparison between the theoretical cross section (black points) and the results of the Apollo run (red dots). Also included is the luminosity of the LHC (blue line).

The agreement is excellent, which is a first confirmation of the proper use of the VEGAS integrator. Also shown is the horizontal exclusion limit as described earlier, at value 3/L, where L=139 fb^{-1}. All cross sections above that line are excluded, which sets the Z' boson mass limit at 5.11 TeV at 95% CL. As expected, this is a better limit than ATLAS 4.5 TeV, given that we assumed a perfect detector, no background and zero events in observation.

ii) Next step is to let the Z' mass float in Apollo. After a proper calibration of the parameter-space normalization, we can collect the generated Z' masses during sampling (and corresponding VEGAS integrals that are used as weights) into a histogram that corresponds to cross section vs. mass. We demonstrate the principle for a short run of ~40M events (that take about a half an hour on a common processor). The cross section function is shown in Fig. 8 (points with errorbars) for a Z' mass range from 4600 GeV to 5400 GeV, compared to dedicated runs with fixed masses (red points).

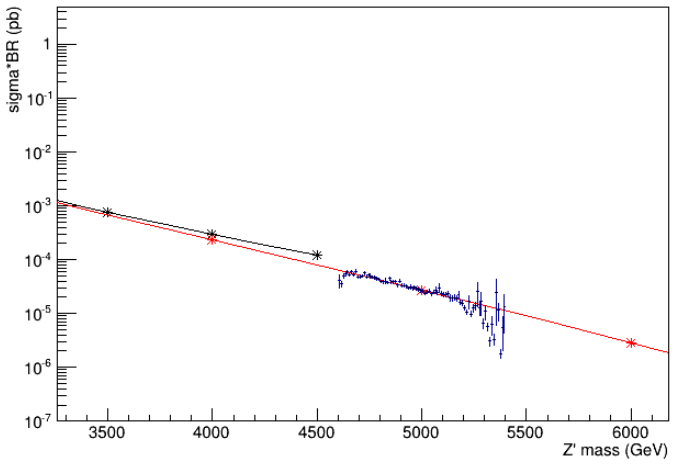


Figure 8 : Integrated cross section (points with errorbars) for $M_{Z'}$ between 4600 and 5400 GeV compared to dedicated runs with fixed masses (red points and line).

We observe a remarkable agreement with expected cross section, which confirms the validity of the method. We see that we can achieve with a single run a better resolution (of 10 GeV), which could be achieved by 100 dedicated simulation jobs in this mass range. Obviously, the distribution suffers from low statistics at higher masses. The solution is a smaller mass range or a longer run to achieve higher statistics. We can fit a range of the high-resolution distribution and compare it with the dedicated runs. From the plot of Fig. 9 we see that they consistently overlap.

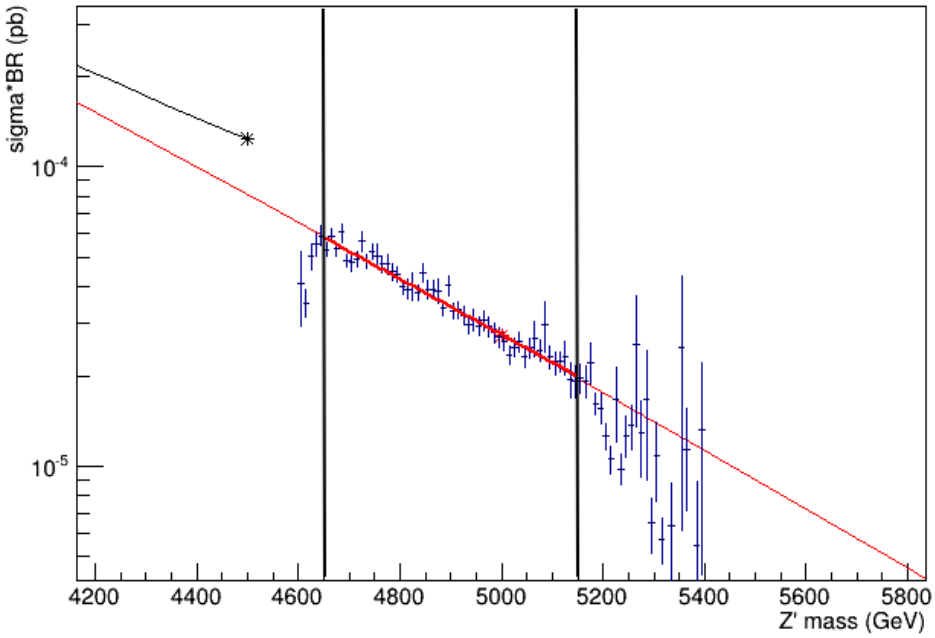


Figure 9 : Fit of the Apollo integrated cross section and comparison to the dedicated runs. The fit uses Apollo sigma in the Z' mass range from 4650 GeV to 5150 GeV. The fit line overlaps with the red line that connected dedicated runs.

iii) Next step is to run Apollo with different masses and coupling factors, which are fixed per job. Now we can create a grid of cross sections for 11 different values of masses (1000 to 6000 GeV with a step of 500 GeV) and 19 different values of coupling factors (0.1 – 1 with a step of 0.1 and 2 – 10 with a step of 1). The 209 combinations are run for few events; since statistics is not an issue here (we get the final cross section answer for each combination of parameter with no need to sample masses and coupling factors yet). We can check which cross sections are above the exclusion limit, which means that the corresponding parameters are excluded. This way we produce the exclusion limit shown in Fig. 10 for a smaller range of coupling factors.

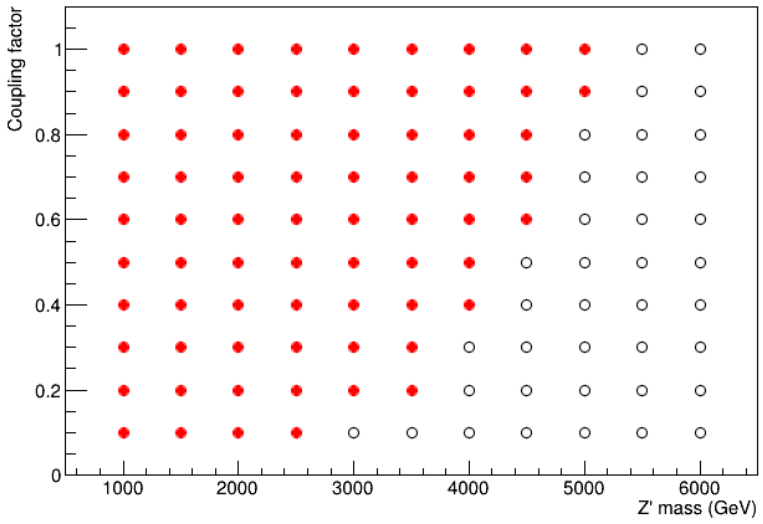


Figure 10 : Repetitive MC generation with fixed parameter values to discern exclusion points, shown in filled red color.

The filled circles show parameter combinations that are excluded and the open circles the parameter combinations that are allowed. This exercise emulates the situation of repetition of MC generation too many times with fixed parameter values each time (a problem that Apollo tries to solve). Eventually we would like to produce such an exclusion limit that it is more efficiently generated and with better resolution in excluded parameter combinations.

iv) Finally, we let both mass and coupling factor float during Apollo integration. This results into a 6-dimensional integration (and corresponding kinematics and parameter generation). We generate 25M events in this multidimensional space, with Z' mass from 4500 and 5500 GeV and a coupling factor range from 0.1 to 1.1. The 2D distribution of the generated parameters can be seen in Fig. 11. Because we adaptively generate parameters based on importance sampling, we generate more events where the cross section is higher. This means higher coupling factors and lower Z' masses.

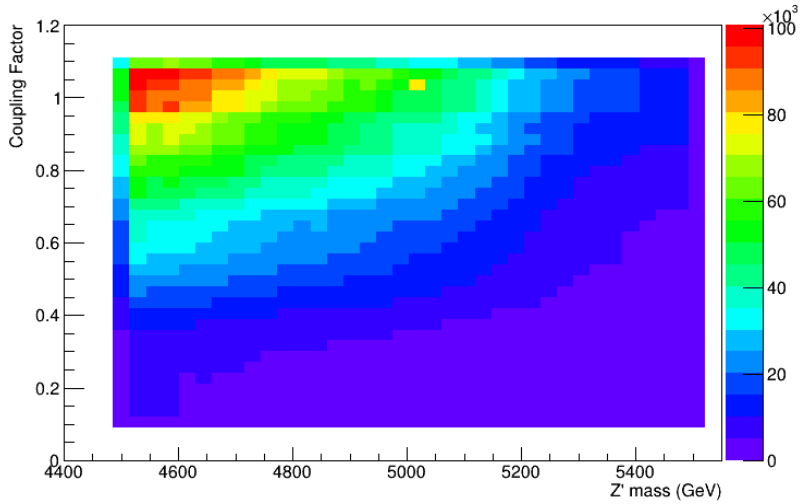


Figure 11 : Sampling of $M_{Z'}$ and coupling factor in the 6-dimensional integration, for $M_{Z'}$ between 4500 and 5500 GeV and a coupling factor between 0.1 and 1.1.

To confirm that we can properly determine the cross sections in the 2-parameter integration (6-dimensional space), we limit the coupling factor around 1 (the SM value), namely from 0.95 to 1.05, and we plot the sum of produced integrals (total cross section within bins) as a function of the mass. Practically speaking, this is a 1D histogram of Z' mass, weighted with VEGAS integral, for this particular range of coupling factors. We see in Fig. 12 that we get back the expected cross section of Z' boson production with SM couplings, as determined by dedicated fixed-parameter runs. The high-mass statistical noise remains, as expected.

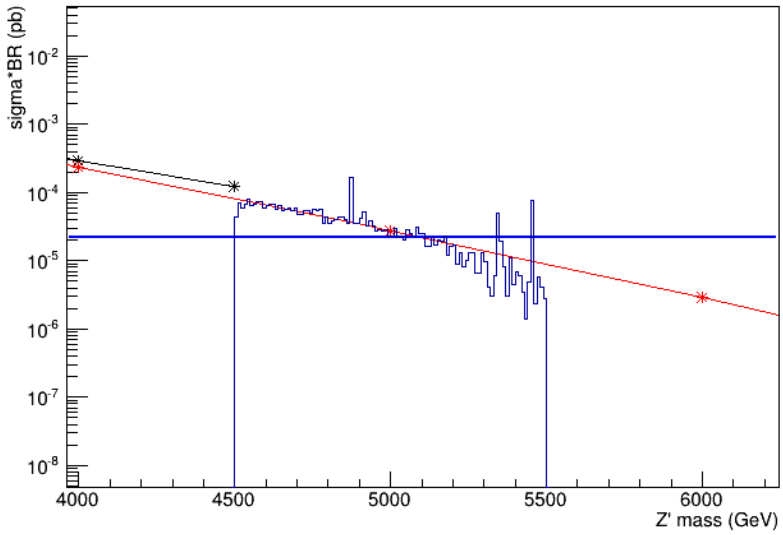


Figure 12 : Produced Apollo cross section limited in a SM-consistent coupling factor, as a function of $M_{Z'}$ (histogram) and comparison to dedicated runs with fixed parameters (red points and line).

The ultimate confirmation of the functionality of Apollo is the production of a 2D exclusion limit using a single run, with a resolution that is better than that of 209 dedicated runs. We demonstrate this for the mass range from 3000 GeV to 5500 GeV and coupling factors from 0.1 to 1.1 (the reason being that these ranges include the boundary between exclusion and not exclusion). We see in Fig. 13 that indeed we cover successfully the expected exclusion area, offering an improved resolution of parameter space. Again, we observe statistical noise, but this is due to the very small number of events (25 M generated per run, for three runs in three mass ranges).

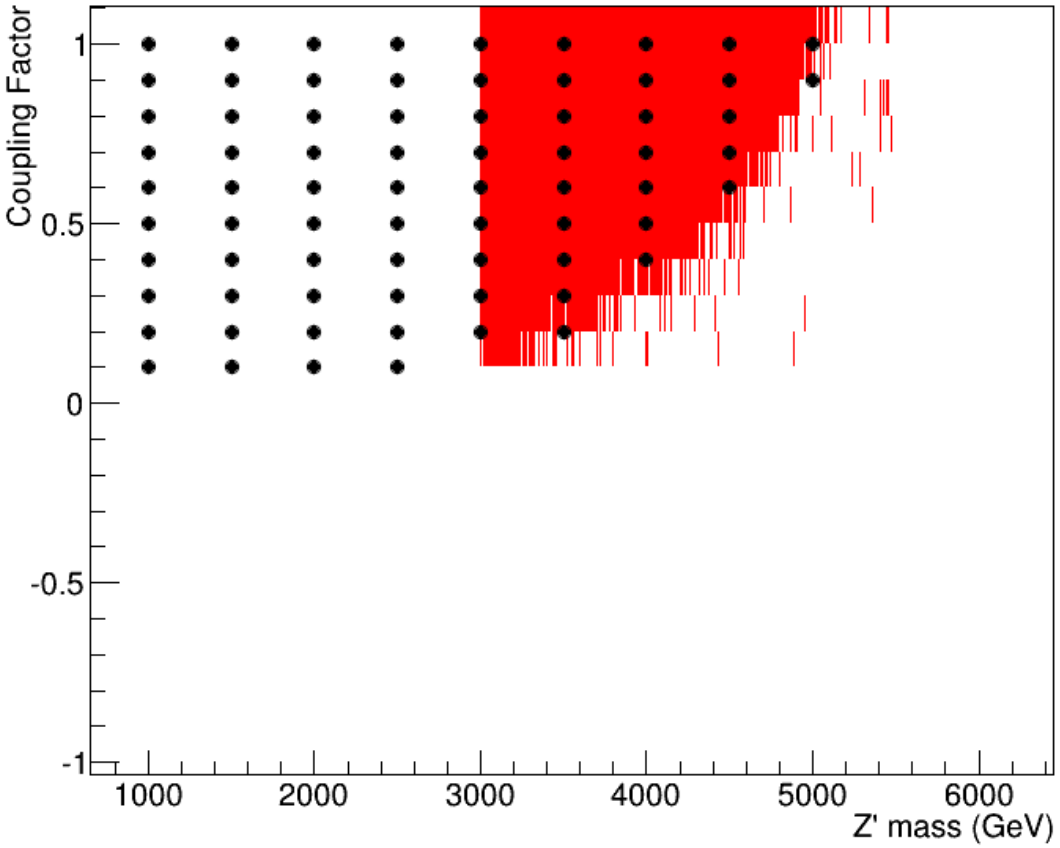


Figure 13 : 2D exclusion limit with $M_{Z'}$ between 3000 and 5500 and coupling factor between 0.1 and 1.1 (red area) as compared to exclusion points from dedicated fixed-parameter runs (black points).

Finally, we can demonstrate the use of Apollo in case of discovery. If a cross section of a new process is measured at 0.022 ± 0.006 fb, we can determine which combinations of Z' masses and coupling factors are consistent with this measurement. The answer is the strip shown in Fig. 14. The content of the contours corresponds to Z' cross sections. This functionality, as well as the exclusion limits or the selection optimization are extremely useful in theories with a larger number of free parameters, such as Supersymmetry.

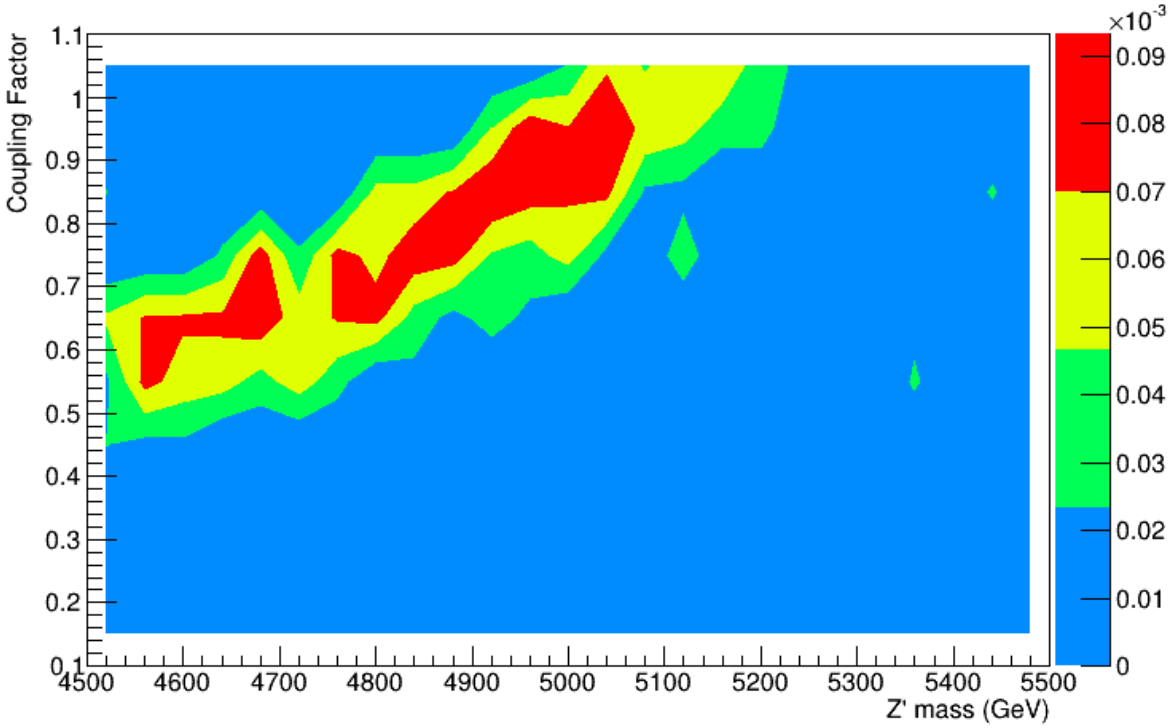


Figure 14 : Combinations of $M_{Z'}$ and coupling factor consistent with a hypothetical discovery with a measured cross section of 0.022 ± 0.006 fb (contours content is proportional to Z' cross section).

Chapter 4: Conclusions

We demonstrated that Apollo can predict correctly the cross sections of a new physics process in a multi-parameter space by adaptively integrating both the kinematic variables and theory parameters. This allows us to efficiently get exclusion limits of these parameters with a single run. At the same time, in case of discovery, it can determine the combinations of parameters that are consistent with the measured cross section. Because of the direct connection of kinematics to a multidimensional parameter space, the program can be used for optimizing the event selection for maximization of signal over background. The value of Apollo is more significant for theories with a large number of free parameters, for example, supersymmetry. In this thesis we demonstrated the principle for a simpler problem of two free parameters in the case of Z' boson production and decay to leptons. In this work we generated only a limited number of events, with the result of statistical noise. Future improvements will include the generation of a larger number of events by running on a computer grid. At the same time, it would be very interesting to utilize Apollo on a SUSY project.

Bibliography

[1] D. Griffiths. *Introduction to Elementary Particles*. Physics textbook. Wiley, 2008.

[2] Francis Halzen and Alan D Martin*. Quark & Leptons: An Introductory Course In Modern Particle Physics*. John Wiley & Sons, 2008.

[3] Donald H. Perkins. *Introdution to High Energy physics 4th edition*. Cambridge University Press, 2000.

[4] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to Quantum Field Theory.* CRC Press, 2018.

[5] The CMS Collaboration *JINST* **3** S08004 (2008).

[6] Lyndon Evans and Philip Bryant *JINST* **3** S08001 (2008).

[7] Jen-Chieh Peng and Jian-Wei Qiu. *The Drell-Yan Process.* The Universe, Vol. 4, No. 3, July-September 2016.

[8] Drell, S.D.; Yan, T.-M. (1970). "Massive Lepton-Pair Production in Hadron-Hadron Collisions at High Energies". Physical Review Letters. **25** (5): 316–320.

[9] G.P. Lepage. *A New Algorithm for Adaptive Multidimensional Integration.* Journal of computational physics **27**, 192-203, 1978.

[10] William H. Press, Brian P Flannery, Saul A. Teukolsky, William T. Vetterling. *Numerical Recipies in FORTRAN 77: The art of scientific computing.* Cambridge University Press. February 1993.

[11] G.P. Lepage. *VEGAS Documentation Release 2.1.4*. February 05 2014.

[12] Alwall, J., Frederix, R., Frixione, S. et al. *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*. J. High Energ. Phys. **2014**, 79 (2014).

[13] Global QCD Analysis of Parton Structure of the Nucleon: CTEQ5 Parton Distributions, European Physics Journal C. 1999.

[14] Rene Brun and Fons Rademakers, ROOT - An Object Oriented Data Analysis Framework, Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996, Nucl. Inst. & Meth. in Phys. Res. A 389 (1997) 81-86.

[15] J. Strologas, “An efficient multi-parameter new-physics event generator”, Pheno 08 Symposium, Apr 28-30, Madison, WI (2008).

[16] ATLAS Collaboration, “Search for new high-mass phenomena in the dilepton final state using 36 fb^{-1} of proton-proton collision data at √s = 13 TeV with the ATLAS detector”, JHEP **10** (2017) 182.

Appendix A

Apollo Structure

Apollo [15] is a program which is the implementation of the idea that we can utilize VEGAS not only to integrate over the kinematic phase but also over free parameters of a theory, thus making Apollo a new-physics particle generator. Applications include setting limits to theories in the absence of discovery or, in case of discovery, a quick determination of the possible combination of values of the free parameters that are consistent with the observed new-physics signal. Apollo is also invaluable at optimizing event selection for maximization of signal over background, because it provides a direct connection of kinematics with the multi-parameter theory space.

The program is made up of four parts (with functions and subroutines incorporated in them as well).

Superscan.f

This is the main body of the program. In this program we call all the necessary subroutines and set the main parameters of our computations such as the number of dimensions and the regions of integration.

We then call the VEGAS function twice, once with a small number of iterations (for example 300) for training. In this stage we don't write the results in our output, this is done to allow VEGAS to adjust to the integrand, by using its sampling techniques, before we actually start writing the sampled momenta and parameter results to a file.

Next, VEGAS is called for the second time with a larger number of iterations (for example 2000). During this integration, we write the sampled kinematics and parameters to a file. The number of training and generation iterations depends on the dimensions of the problem (i.e. number of outgoing particles and number of theory parameters).

Function Hadronic Integrand:

Here we call the particle distribution functions of each quark from the pdf library SetCtq5. Then, by calling the fxn function below for each process and summing over all the 5 quark pairs, we get the final Hadronic Integrand that will be the function integrated by VEGAS in our main program.

fxn.f

Function that calculates the partonic cross section for each of the 10 given interactions between the five quarks up, down, strange, charmed and beauty (not the top quark as it is too heavy to be found within the proton) and its corresponding antiquark. The number of interactions is double the number of quarks because of the combinatorics: each quark can originate from either the colliding protons.

The partonic cross section is a product of the Madgraph-generated matrix element of each interaction times a kinematic phase-space factor times a factor for the conversion of results to pb.

vegas.f

This file includes the VEGAS subroutine whose algorithm has already been presented. The actual implementation can be found in Appendix B. It is called by superscan.f which also sets its integration parameters (number of dimensions, region) and is given the integral from the Hadronic_Integral subroutine also included in superscan.f.

Here is where we receive our final output, a data file which includes the energy and 3-dimensional momenta of every particle, the sampled parameters of the theory, and the final integral which gives us the cross section after certain corrections.

Matrix_Element.f

These are several separate files that are produced by Madgraph in FORTRAN with the use of the helicity package HELAS. They are called by the fxn function individually to contribute to the overall partonic cross section. Each file contains the matrix element of one of the following partonic interactions:

$$u\bar{u}, \bar{u}u, d\bar{d}, \bar{d}d, c\bar{c}, \bar{c}c, s\bar{s}, \bar{s}s, b\bar{b}, \bar{b}b$$

These are all the partonic interactions that lead to the Drell-Yan interaction under investigation. An example of the matrix-element code, modified for Z' production, can be seen in Appendix C.

Appendix B

VEGAS implementation

We present here the VEGAS algorithm as implemented in Fortran 77 in the Numerical Recipes [10].

```
#include <stdio.h>
#include <math.h>

#include "nrutil.h"
#define ALPH 1.5
#define NDMX 50

#define MXDIM 10
#define TINY 1.0e-30
extern long idum; !For random number initialization in main.
void vegas(float regn[], int ndim, float (*fxn)(float [], float), int init,
unsigned long ncall, int itmx, int nprn, float *tgral, float *sd,
float *chi2a)
!Performs Monte Carlo integration of a user-supplied ndim-dimensional function fxn over a
!rectangular volume specified by regn[1..2*ndim], a vector consisting of ndim "lower left"
!coordinates of the region followed by ndim "upper right" coordinates. The integration consists
!of itmx iterations, each with approximately ncall calls to the function. After each iteration
!the grid is refined; more than 5 or 10 iterations are rarely useful. The input flag init signals
!whether this call is a new start, or a subsequent call for additional iterations (see comments
!below). The input flag nprn (normally 0) controls the amount of diagnostic output. Returned
!answers are tgral (the best estimate of the integral), sd (its standard deviation), and chi2a(χ2
!per degree of freedom, an indicator of whether consistent results are being obtained). See
!text for further details.
{
float ran2(long *idum);
void rebin(float rc, int nd, float r[], float xin[], float xi[]);
static int i,it,j,k,mds,nd,ndo,ng,npg,ia[MXDIM+1],kg[MXDIM+1];
static float calls,dv2g,dxg,f,f2,f2b,fb,rc,ti,tsi,wgt,xjac,xn,xnd,xo;
static float d[NDMX+1][MXDIM+1],di[NDMX+1][MXDIM+1],dt[MXDIM+1],
dx[MXDIM+1], r[NDMX+1],x[MXDIM+1],xi[MXDIM+1][NDMX+1],xin[NDMX+1];
static double schi,si,swgt;
!Best make everything static, allowing restarts.
if (init <= 0) { Normal entry. Enter here on a cold start.
mds=ndo=1; !Change to mds=0 to disable stratified sampling,
i!.e., use importance sampling only.
for (j=1;j<=ndim;j++) xi[j][1]=1.0;
}
if (init <= 1) si=swgt=schi=0.0;
Enter here to inherit the grid from a previous call, but not its answers.

if (init <= 2) { !Enter here to inherit the previous grid and its answers.
nd=NDMX;
ng=1;
if (mds) { !Set up for stratification.
ng=(int)pow(ncall/2.0+0.25,1.0/ndim);

mds=1;
if ((2*ng-NDMX) >= 0) {
mds = -1;
```

```
npg=ng/NDMX+1;
nd=ng/npg;
ng=npg*nd;
}
}
for (k=1,i=1;i<=ndim;i++) k *= ng;
npg=IMAX(ncall/k,2);
calls=(float)npg * (float)k;

dxg=1.0/ng;
for (dv2g=1,i=1;i<=ndim;i++) dv2g *= dxg;
dv2g=SQR(calls*dv2g)/npg/npg/(npg-1.0);

xnd=nd;
dxg *= xnd;
xjac=1.0/calls;

for (j=1;j<=ndim;j++) {
dx[j]=regn[j+ndim]-regn[j];
xjac *= dx[j];
}
if (nd != ndo) { !Do binning if necessary.
for (i=1;i<=IMAX(nd,ndo);i++) r[i]=1.0;
for (j=1;j<=ndim;j++) rebin(ndo/xnd,nd,r,xin,xi[j]);
ndo=nd;
}
if (nprn >= 0) {
printf("%s: ndim= %3d ncall= %8.0f\n",
" Input parameters for vegas",ndim,calls);
printf("%28s it=%5d itmx=%5d\n"," ",it,itmx);
printf("%28s nprn=%3d ALPH=%5.2f\n"," ",nprn,ALPH);

printf("%28s mds=%3d nd=%4d\n"," ",mds,nd);
for (j=1;j<=ndim;j++) {
printf("%30s xl[%2d]= %11.4g xu[%2d]= %11.4g\n",
" ",j,regn[j],j,regn[j+ndim]);
}
}
}
for (it=1;it<=itmx;it++) {
```

!Main iteration loop. Can enter here (`init` ≥ 3) to do an additional `itmx` iterations with all other !parameters unchanged.

```
ti=tsi=0.0;

for (j=1;j<=ndim;j++) {
kg[j]=1;
for (i=1;i<=nd;i++) d[i][j]=di[i][j]=0.0;
}
for (;;) {
fb=f2b=0.0;
for (k=1;k<=npg;k++) {
wgt=xjac;

for (j=1;j<=ndim;j++) {
xn=(kg[j]-ran2(&idum))*dxg+1.0;
ia[j]=IMAX(IMIN((int)(xn),NDMX),1);
```

```
if (ia[j] > 1) {
xo=xi[j][ia[j]]-xi[j][ia[j]-1];
rc=xi[j][ia[j]-1]+(xn-ia[j])*xo;
} else {
xo=xi[j][ia[j]];
rc=(xn-ia[j])*xo;
}
x[j]=regn[j]+rc*dx[j];
wgt *= xo*xnd;
}
f=wgt*(*fxn)(x,wgt);

f2=f*f;
fb += f;
f2b += f2;

for (j=1;j<=ndim;j++) {
di[ia[j]][j] += f;
if (mds >= 0) d[ia[j]][j] += f2;
}
}
f2b=sqrt(f2b*npg);
f2b=(f2b-fb)*(f2b+fb);
if (f2b <= 0.0) f2b=TINY;

ti += fb;
tsi += f2b;
if (mds < 0) { !Use stratified sampling.
for (j=1;j<=ndim;j++) d[ia[j]][j] += f2b;
}
for (k=ndim;k>=1;k--) {
kg[k] %= ng;
if (++kg[k] != 1) break;
}

if (k < 1) break;
}
tsi *= dv2g; !Compute final results for this iteration.
wgt=1.0/tsi;
si += wgt*ti;

schi += wgt*ti*ti;
swgt += wgt;
*tgral=si/swgt;

*chi2a=(schi-si*(*tgral))/(it-0.9999);
if (*chi2a < 0.0) *chi2a = 0.0;
*sd=sqrt(1.0/swgt);

tsi=sqrt(tsi);
if (nprn >= 0) {
printf("%s %3d : integral = %14.7g +/- %9.2g\n",
" iteration no.",it,ti,tsi);
printf("%s integral =%14.7g+/-%9.2g chi**2/IT n = %9.2g\n",
" all iterations: ",*tgral,*sd,*chi2a);
if (nprn) {
for (j=1;j<=ndim;j++) {
printf(" DATA FOR axis %2d\n",j);
printf("%6s%13s%11s%13s%11s%13s\n",
```

```
"X","delta i","X","delta i","X","delta i");
for (i=1+nprn/2;i<=nd;i += nprn+2) {
printf("%8.5f%12.4g%12.5f%12.4g%12.5f%12.4g\n",
xi[j][i],di[i][j],xi[j][i+1],
di[i+1][j],xi[j][i+2],di[i+2][j]);
}
}
}
}

for (j=1;j<=ndim;j++) {
```
!Refine the grid. Consult references to understand the subtlety of this procedure. The refinement is !damped, to avoid rapid, destabilizing changes, and also compressed in range by the exponent `ALPH`.
```
xo=d[1][j];
xn=d[2][j];

d[1][j]=(xo+xn)/2.0;
dt[j]=d[1][j];
for (i=2;i<nd;i++) {
rc=xo+xn;
xo=xn;
xn=d[i+1][j];

d[i][j] = (rc+xn)/3.0;
dt[j] += d[i][j];
}
d[nd][j]=(xo+xn)/2.0;
dt[j] += d[nd][j];
}
for (j=1;j<=ndim;j++) {
rc=0.0;

for (i=1;i<=nd;i++) {
if (d[i][j] < TINY) d[i][j]=TINY;
r[i]=pow((1.0-d[i][j]/dt[j])/
(log(dt[j])-log(d[i][j])),ALPH);
rc += r[i];
}
rebin(rc/xnd,nd,r,xin,xi[j]);
}
}
}
void rebin(float rc, int nd, float r[], float xin[], float xi[])
```
!Utility routine used by `vegas`, to rebin a vector of densities `xi` into new bins defined by a vector `r`.
```
{
int i,k=0;
float dr=0.0,xn=0.0,xo=0.0;
for (i=1;i<nd;i++) {
while (rc > dr)
dr += r[++k];
if (k > 1) xo=xi[k-1];
xn=xi[k];
dr -= rc;
xin[i]=xn-(xn-xo)*dr/r[k];
}
for (i=1;i<nd;i++) xi[i]=xin[i];

xi[nd]=1.0;
}
```

Appendix C

Modified matrix element for Z' boson study

We present here an example of a matrix element as generated by Madgraph, with our Z'-boson alterations shown in red. This is the matrix element for the process $u + \bar{u} \rightarrow Z' \rightarrow e^+ e^-$

```
      SUBROUTINE SUUB_EPEM(P1,ANS, MZ_, couplingFactor)
C
C FUNCTION GENERATED BY MADGRAPH
C RETURNS AMPLITUDE SQUARED SUMMED/AVG OVER COLORS
C AND HELICITIES
C FOR THE POINT IN PHASE SPACE P(0:3,NEXTERNAL)
C
C FOR PROCESS : u u~ -> e+ e-
C
C Crossing   1 is u u~ -> e+ e-
      IMPLICIT NONE
C
C CONSTANTS
C
      INTEGER    NEXTERNAL,   NCOMB,     NCROSS
      PARAMETER (NEXTERNAL=4, NCOMB= 16, NCROSS=  1)
      INTEGER    THEL
      PARAMETER (THEL=NCOMB*NCROSS)
C
C ARGUMENTS
C
      REAL*8 P1(0:3,NEXTERNAL),ANS(NCROSS)
      REAL*8 MZ_, couplingFactor
C
C LOCAL VARIABLES
C
      INTEGER NHEL(NEXTERNAL,NCOMB),NTRY
      REAL*8 T, P(0:3,NEXTERNAL)
      REAL*8 UUB_EPEM
      INTEGER IHEL,IDEN(NCROSS),IC(NEXTERNAL,NCROSS)
      INTEGER IPROC,JC(NEXTERNAL)
      LOGICAL GOODHEL(NCOMB,NCROSS)
      DATA GOODHEL/THEL*.FALSE./
      DATA NTRY/0/
      DATA (NHEL(IHEL,  1),IHEL=1,4) / -1, -1, -1, -1/
      DATA (NHEL(IHEL,  2),IHEL=1,4) / -1, -1, -1,  1/
      DATA (NHEL(IHEL,  3),IHEL=1,4) / -1, -1,  1, -1/
      DATA (NHEL(IHEL,  4),IHEL=1,4) / -1, -1,  1,  1/
      DATA (NHEL(IHEL,  5),IHEL=1,4) / -1,  1, -1, -1/
      DATA (NHEL(IHEL,  6),IHEL=1,4) / -1,  1, -1,  1/
      DATA (NHEL(IHEL,  7),IHEL=1,4) / -1,  1,  1, -1/
      DATA (NHEL(IHEL,  8),IHEL=1,4) / -1,  1,  1,  1/
```

```
      DATA (NHEL(IHEL,  9),IHEL=1,4) /  1, -1, -1, -1/
      DATA (NHEL(IHEL, 10),IHEL=1,4) /  1, -1, -1,  1/
      DATA (NHEL(IHEL, 11),IHEL=1,4) /  1, -1,  1, -1/
      DATA (NHEL(IHEL, 12),IHEL=1,4) /  1, -1,  1,  1/
      DATA (NHEL(IHEL, 13),IHEL=1,4) /  1,  1, -1, -1/
      DATA (NHEL(IHEL, 14),IHEL=1,4) /  1,  1, -1,  1/
      DATA (NHEL(IHEL, 15),IHEL=1,4) /  1,  1,  1, -1/
      DATA (NHEL(IHEL, 16),IHEL=1,4) /  1,  1,  1,  1/
      DATA (  IC(IHEL,  1),IHEL=1,4) /  1,  2,  3,  4/
      DATA (IDEN(IHEL),IHEL=  1,  1) /  36/
C ----------
C BEGIN CODE
C ----------
      NTRY=NTRY+1
      DO IPROC=1,NCROSS
      CALL SWITCHMOM(P1,P,IC(1,IPROC),JC,NEXTERNAL)
      DO IHEL=1,NEXTERNAL
         JC(IHEL) = +1
      ENDDO
      ANS(IPROC) = 0D0
      DO IHEL=1,NCOMB
          IF (GOODHEL(IHEL,IPROC) .OR. NTRY .LT. 2) THEN
             T=UUB_EPEM(P ,NHEL(1,IHEL),JC(1), MZ_, couplingFactor)
             ANS(IPROC)=ANS(IPROC)+T
              IF (T .GT. 0D0 .AND. .NOT. GOODHEL(IHEL,IPROC)) THEN
                  GOODHEL(IHEL,IPROC)=.TRUE.
C             WRITE(*,*) IHEL,T
              ENDIF
          ENDIF
      ENDDO
      ANS(IPROC)=ANS(IPROC)/DBLE(IDEN(IPROC))
      ENDDO
      END

  REAL*8 FUNCTION UUB_EPEM(P,NHEL,IC, MZ_, couplingFactor)
C
C FUNCTION GENERATED BY MADGRAPH
C RETURNS AMPLITUDE SQUARED SUMMED/AVG OVER COLORS
C FOR THE POINT WITH EXTERNAL LINES W(0:6,NEXTERNAL)
C
C FOR PROCESS : u u~ -> e+ e-
C
      IMPLICIT NONE
C
C CONSTANTS
C
      INTEGER    NGRAPHS,    NEIGEN,     NEXTERNAL
      PARAMETER (NGRAPHS=   2,NEIGEN=  1,NEXTERNAL=4)
      INTEGER    NWAVEFUNCS     , NCOLOR
      PARAMETER (NWAVEFUNCS=   6, NCOLOR=   1)
      REAL*8     ZERO
      PARAMETER (ZERO=0D0)
C
```

```
C ARGUMENTS
C
      REAL*8 P(0:3,NEXTERNAL)
      INTEGER NHEL(NEXTERNAL), IC(NEXTERNAL)
      REAL*8 MZ_, couplingFactor
C
C LOCAL VARIABLES
C
      INTEGER I,J
      COMPLEX*16 ZTEMP
      REAL*8 DENOM(NCOLOR), CF(NCOLOR,NCOLOR)
      COMPLEX*16 AMP(NGRAPHS), JAMP(NCOLOR)
      COMPLEX*16 W(6,NWAVEFUNCS)
      REAL*8 couplingLeft, couplingRight
C
C GLOBAL VARIABLES
C
      INCLUDE 'coupl.inc'

C
C COLOR DATA
C
      DATA Denom(1  )/            1/
      DATA (CF(i,1  ),i=1  ,1  ) /     3/
C               T[2,1]
C ----------
C BEGIN CODE
C ----------
      CALL IXXXXX(P(0,1   ),ZERO ,NHEL(1   ),+1*IC(1   ),W(1,1   ))
      CALL OXXXXX(P(0,2   ),ZERO ,NHEL(2   ),-1*IC(2   ),W(1,2   ))
      CALL IXXXXX(P(0,3   ),ZERO ,NHEL(3   ),-1*IC(3   ),W(1,3   ))
      CALL OXXXXX(P(0,4   ),ZERO ,NHEL(4   ),+1*IC(4   ),W(1,4   ))
      CALL JIOXXX(W(1,1   ),W(1,2   ),GAU ,ZERO    ,ZERO    ,W(1,5   ))
      CALL IOVXXX(W(1,3   ),W(1,4   ),W(1,5   ),GAL ,AMP(1   ))
      CALL JIOXXX(W(1,1   ),W(1,2   ),GZU ,MZ_      ,ZWIDTH      ,W(1,6))

      couplingLeft = 0.1973*couplingFactor
      couplingRight = -0.173*couplingFactor
      gzl(1)= couplingLeft
      gzl(2)= couplingRight

      CALL IOVXXX(W(1,3   ),W(1,4   ),W(1,6   ),GZL ,AMP(2   ))
      JAMP(   1) = +AMP(   1)+AMP(   2)
      UUB_EPEM = 0.D0
      DO I = 1, NCOLOR
          ZTEMP = (0.D0,0.D0)
          DO J = 1, NCOLOR
              ZTEMP = ZTEMP + CF(J,I)*JAMP(J)
          ENDDO
          UUB_EPEM =UUB_EPEM+ZTEMP*DCONJG(JAMP(I))/DENOM(I)
      ENDDO
C      CALL GAUGECHECK(JAMP,ZTEMP,EIGEN_VEC,EIGEN_VAL,NCOLOR,NEIGEN)
      END
```