

Search for Weak Scale Supersymmetric Particles in Compressed Scenarios

©2023

Justin Anguiano

B.S. Engineering Physics, University of Kansas, 20XX

M.S. Computational Physics and Astronomy, University of Kansas, 20XX

Submitted to the graduate degree program in Department of Physics and Astronomy and
the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements
for the degree of Doctor of Philosophy.

Graham Wilson, Chairperson

Alice Bean, Co-Chair

Committee members

Christopher Rogan

Ian Lewis

Zsolt Talata, External Reviewer

Date defended: _____ July 02, 2019 _____

The Dissertation Committee for Justin Anguiano certifies
that this is the approved version of the following dissertation :

Search for Weak Scale Supersymmetric Particles in Compressed Scenarios

Graham Wilson, Chairperson

Date approved: _____ August 06, 2019 _____

Abstract

This is the abstract

Acknowledgements

Thanks everybody

Contents

1 Analysis Description	1
1.1 Introduction and Strategy	1
1.2 Data and Simulation	1
1.3 Event Selection and Physics Objects	2
1.4 Categorization and Fit Strategy	6

List of Figures

1.1	grids	3
1.2	met trig sf	4
1.3	cleaning cuts	5

List of Tables

1.1	table caption	2
1.2	table caption2	6

Chapter 1

The Standard Model and Supersymmetry

1.1 Introduction

The fundamental building blocks of matter and their interactions expressed through three of the four fundamental forces of nature via the Standard Model (SM). The fourth force, or gravity, is left to General Relativity. The SM is the culmination of over a century of work by many scientists, and has its roots in the late 19th century. The first force is based on the theory of strong interactions, formulated as quantum chromodynamics by Murray Gell-Mann and others in the 1960s, which provides a description of how protons and neutrons are held together in the nucleus of an atom. The theory describing weak interactions was developed by Enrico Fermi in the 1930s, which was then combined with electromagnetic interactions in electroweak theory by Sheldon Glashow, Abdus Salam, and Steven Weinberg in the 1960s. In this chapter, concepts of the Standard Model are introduced, including the fundamental particles, fields, and their basic properties and interactions. Then expanding from the core SM we will discuss an extension of the Standard Model with supersymmetry, which proposes a new symmetry between fermions and bosons. Finally, we delve into the specifics of simplified models of supersymmetry and the challenges associated with detecting these models experimentally.

1.2 The Standard Model

The Standard Model is a collection of adhoc theories used to predict and reproduce experimental data. The theory itself incorporates four major concepts: Quantum Field theory (QFT), the Dirac equation, the gauge principle, and the Higgs mechanism. These four principles are constrained by physical data and describe the set of elementary particles, known as fermions and bosons. The SM generally refers to the SM Lagrangian, an equation with different sectors that describe different subsets of particles, fields, and their interactions. The SM Lagrangian itself consists of 26 free parameters which are input by hand. These parameters are: the masses of the 12 fermions, 3 coupling constants that describe gauge interactions: g, g', g_s , 2 parameters to describe the Higgs potential i.e. the higgs mass m_h and the vacuum expectation value (vev), and 9 mixing angles which describe the PMNS and CKM matrices or the mixing of different fermionic fields. The 12 fermion paramters are subdivided by three neutrinos m_{ν_i} , three charged leptons $m_{\ell_i}^\pm$, and six quarks m_{q_i} ;

QFT provides a description for both known and theoretical particles by combining quantum theory, the field concept, and relativity (cite peskin). The gauge theory aspect describes the exact nature of QFT interactions and provides the mechanisms for the electromagnetic, strong, and weak forces. We know of three gauge fields: \vec{G} which transforms under $SU(3)$ and govern strong interactions, \vec{W} and B which transform under $SU(2)_L \times U(1)$ and govern electromagnetic and weak interactions. The combination of the gauge fields and fermion fields along with the Dirac equation yields eigenstates that represent fermionic matter particles. These particles would be massless if not for the inclusion of the complex scalar Higgs field. The spontaneous symmetry breaking of the Higgs field, due to the Yukawa coupling, creates the non-zero vev responsible for generating the masses of the electroweak gauge bosons. Additionally, the interaction between the fermionic fields and the non zero-vev generates the masses of SM fermions.

The set of standard model elementary particles is divided into two subgroups: fermions and

bosons. The fermions consist of both charged and neutral leptons as well as fractionally charged quarks. There are three flavors of charged leptons (ℓ), the electron (e), the muon (μ), and the tau (τ). Each charged lepton has a flavor pairing neutral neutrino ν_ℓ . The e and μ are also generally considered as "light" leptons due to their small mass relative to the τ . The term lepton, depending on context, often refers to only the charged particles. As for the quarks, there are also three generations of pairs of quarks.. The highest set of quarks are the up (u) and down (d) quarks, followed by the charm (c) and strange (s), and lastly the bottom (b) and extremely massive top quark (t). The bosons are the force carrying particles which represent the gauge fields. They are comprised of the vector bosons - the photon (γ), gluon (g), the W^\pm , and the Z^0 - along with the singular scalar boson the Higgs (h). The elementary particles masses, generations, and spins are summarized in Figure ??.

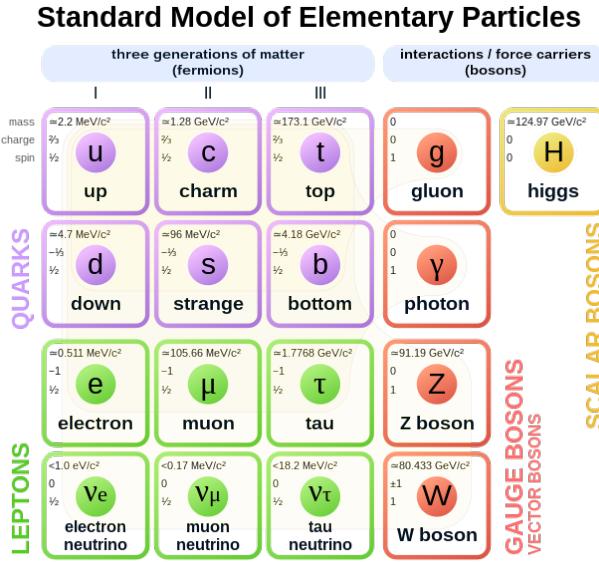


Figure 1.1: particles figure cite wiki

The SM is an asymmetric chiral theory, combining three groups $SU(3)_L \times SU(2)_L \times U(1)$. The L , or left handed, subscript indicates that mirrored fields (with different chiralities) transform differently under the Lorentz group and the EW gauge group (cite slides). The consequence of chirality is that the possible combinations between interaction vertices is limited(cite thompson). This peculiar property shows up with the W boson, which only

couples to left handed particles or right handed antiparticles. Extensions of the standard model also often extend chiral or symmetrical properties.

1.3 Supersymmetry

Supersymmetry (SUSY) is an extension of the standard model. It adds a generator that rotates the spin between bosons and fermions. This then introduces a bosonic degree of freedom for every fermionic degree of freedom (cite run2 susy paper) which generates a super partner for each particle with spin differing by a half integer. The resulting set of mirrored elementary particles are referred to as sparticles. Each bosonic sparticle carries the same name as its fermion partner but with an "s" prefix e.g. sfermion, squark, selectron. As for the bosons, with the gauge fields B and \vec{W} , these are accompanied by three super symmetric fields - the Higgsino \tilde{H} , Bino \tilde{B} , and Wino \tilde{W} . The mixture of the B and \vec{W} SM fields can be represented by particle matrix. One can obtain the mass eigenstates representing the SM particles γ , Z , W^\pm through the diagonalization of particle matrix. Similarly, the Higgsino, Bino, and Wino mix to produce four neutral and two charged eigenstates, the neutralinos ($\tilde{\chi}_1^0, \tilde{\chi}_2^0, \tilde{\chi}_3^0, \tilde{\chi}_4^0$) and charginos ($\tilde{\chi}_1^\pm, \tilde{\chi}_2^\pm$) (cite erich's 43 susy matrix eigenstates). SUSY also requires an additional Higgs doublet to give mass to up-type and down-type fermions, (cite this run 2 paper.. reference chasing) leading to five higgs boson states consisting of two charged Higgs and three neutral Higgs. The lightest neutral higgs of the three neutral options represents the SM Higgs boson. The full set of SM particles alongside their SUSY partners are illustrated in Figure ???. The addition of another higgs doublet also introduces a second vev. The ratio between the two vev's is commonly denoted as $v_1/v_2 = \tan \beta$ and is an important parameter in experimental searches. Another important bookkeeping parameter, similar to lepton number or baryon number conservation, is R-parity. This parameter tallies the total number of SM particles (+1) and sparticles (-1) and expects the net total between particles to be conserved in the initial and final states. R-parity conservation then requires

sparticles to be produced in pairs. If R-parity is violated, the common consequence is that the lightest supersymmetric particle (LSP) is unstable.

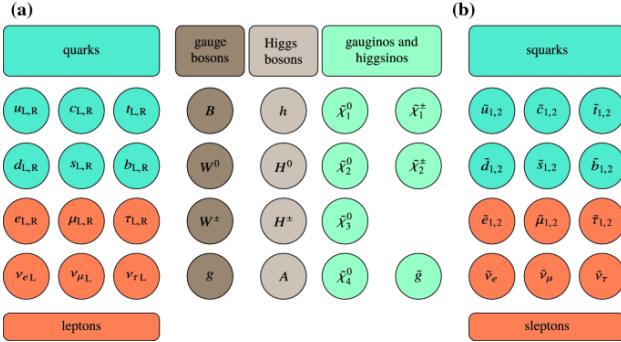


Figure 1.2: stolen from this springer thesis book, probably make my own figure later https://link.springer.com/chapter/10.1007/978-3-030-25988-4_4

Supersymmetry is an extremely expansive model and intractable to experimentally test without significant well motivated simplifications. The most experimentally common simplified SUSY model is the Minimally Super Symmetric Standard Model (MSSM). The MSSM contains the smallest number of new particle states and new interactions which are consistent with phenomenology (cite howie direct weak scale book). The MSSM is still experimentally inaccessible due to the presence of over 100 parameters, where small changes in parameter space can completely morph the model structure and experimental signatures. To reduce the problem's dimensionality, further simplification is needed, resulting in a popular simplified model: the phenomenological MSSM (pMSSM). The pMSSM contains 19 parameters which include the masses of each generation of squark and slepton, parameters to control the mixing of $\tilde{H}, \tilde{W}, \tilde{B}$, and dials for the higgs doublet(cite what wiki cites). The pMSSM is still borderline too complicated to attack directly, so, the pMSSM is boiled down into a simplified model of four parameters M_1, M_2, μ , and $\tan \beta$. M_1 and M_2 are the gaugino mass parameters, μ is the Higgsino mass parameter, and $\tan \beta$ is the previously mentioned vev ratio (cite Fuks paper). A model point from this four parameter space is referred to as Realistic simplified gaugino-higgsino model, and targets specific regions of MSSM parameter space and experimental topologies.

To effectively grasp the structure of SUSY and various models, either in the pMSSM or simplified models, there are a couple key elements to consider. The first elements is the mass scale of the relative SUSY sectors i.e. how massive are the gauginos versus sleptons versus squarks. If the mass scales are well separated, the sectors are effectively decoupled. If the mass scales are similar then it may introduce complicated cross-talk between sectors. In an electroweak SUSY search with a 4 parameter simplified model, the model can be further simplified by assuming squarks and slepton masses sit at the several TeV scale while the targeted electroweak-inos are at detectable TeV and sub-TeV scale. By decoupling sectors outside the sector-of-interest we remove the interaction between these groups, so, if sleptons are decoupled from the gauginos complicated dependencies, like cascading decays are avoided. The other key element is the composition of the LSP, typically $\tilde{\chi}_1^0$. Each unique model point is composed of a specific mixing of $\tilde{H}, \tilde{W}, \tilde{B}$ with an LSP that reflects that mixing. The model point is denoted by the field that dominates the overall mix, so a Higgsino model has an LSP composed of mostly \tilde{H} (cite mixing altas paper?). The characteristic take away from simplified model types is that H,W,B can control the nature of the model by governing the overall cross sections for sparticles, the topological infrastructure, and how the sparticles interact and amongst themselves and SM particles. Two pMSSM examples comparing the mass structure between two arbitrary mass points of a Wino model versus Higgsino model is shown in Figure ???. For both models the Higgs and slepton sectors are decoupled at a multi-TeV scale while the squark and gaugino sectors are at an accesible TeV and sub-TeV scale. Note that small changes in pMSSM model space results in differing LSP content and large variations in the relative mass structure and orderings. The difference in cross sections between the same two model points for gaugino pair production combinations are show in in Figure ???. This relative differences in cross section illustrates that the same tweak in parameter space can induce order of magnitude changes sparticle production.

In addition to the mass structure and cross sections, the decay nature of H/W/B models also varies. The variation in decay modes has a significant impact on the experimental

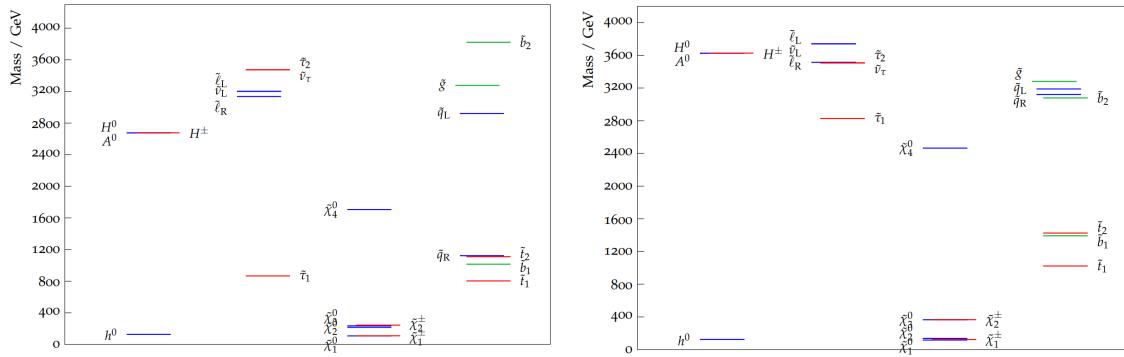


Figure 1.3: mass structure winno vs higgsino modelpoints

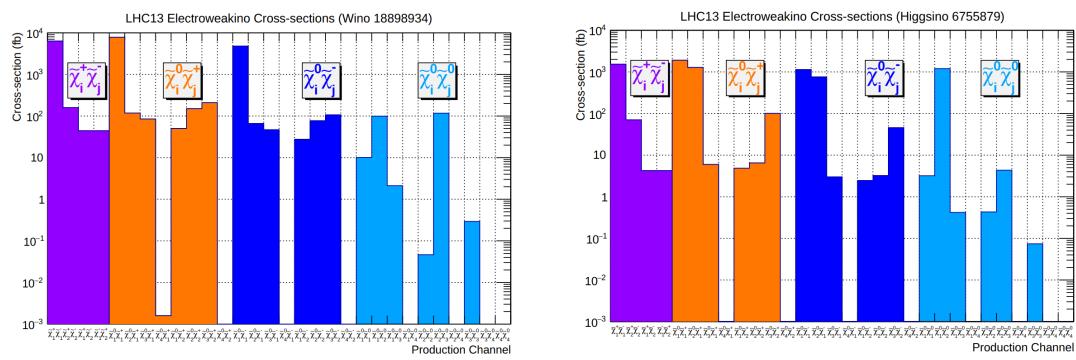


Figure 1.4: xsec strucutre wino vs higgsino modelpoints

channels and signatures of interest. In an experimental search we would expect the heavier sparticles to decay to both SM particles along with the LSP. If the LSP happens to be close in mass to its parent, say $O(100)$ GeV or less, the model would be considered as a compressed scenario. This scenario is considered compressed because the observable energy of the SM particle involved in a sparticle decay is compressed to a very small amount due to the majority of the available energy being used by the rest mass of the sparticles. Of the 3 types of models, the most likely candidates for compression are the Higgsino-like and Bino-like models. Wino models by far have the largest cross sections but are the least likely to have compressed states. Particularly interesting topologies for these compressed models involve decay signatures of processes like $\tilde{\chi}_2^0 \rightarrow Z^* \tilde{\chi}_1^0$, $\tilde{\chi}_2^0 \rightarrow \tilde{\chi}_1^\pm \tilde{\chi}_1^0$, $\tilde{\chi}_1^\pm \rightarrow W^\pm \tilde{\chi}_1^0$, $\tilde{t} \rightarrow t \tilde{\chi}_1^0$, $\tilde{\ell} \rightarrow \ell \tilde{\chi}_1^0$. The nature of sparticle decay is not only dependent on the H/W/B nature of the model but also on the degree of compression. Figure ?? shows the average decay modes for H W or B from a selection of pMSSM models (cite atlas pmssm paper).

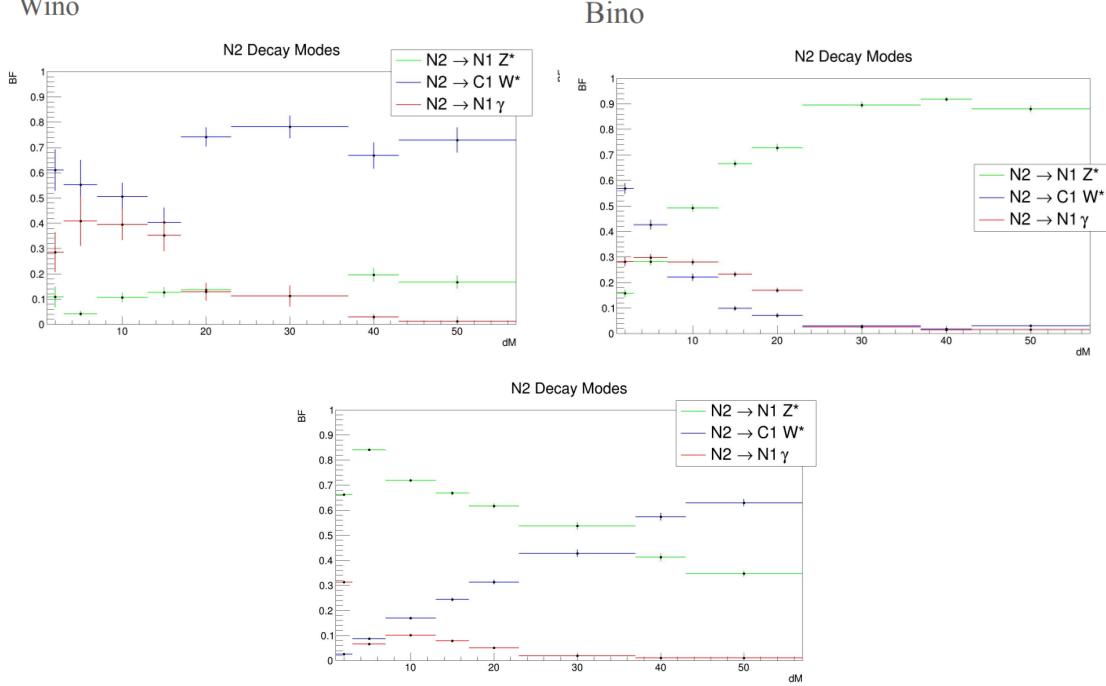


Figure 1.5: N2 BFs

Note that between each model type in Figure ?? the Z^* and W^\pm modes can be highly sup-

pressed or enhanced. In some cases even, specific modes like $\tilde{\chi}_2^0 \rightarrow \tilde{\chi}_1^\pm W^\mp$ can be either kinematically forbidden, or excluded to streamline MC production and enhance the statistical power of different targeted final states. Alongside the decay specific complications, the phase space of the final state particles is model dependent. For instance, in the case of $\tilde{\chi}_2^0 \rightarrow Z^* \tilde{\chi}_1^0$ the shape of Z dilepton mass distribution $m_{\ell\ell}$ changes depending on the sign of the gaugino eigenstates. Experimentally this problem is divided into two possible scenarios: cases where the eigenstates are the same sign and cases where the eigenstates are the opposite sign. The distribution that showcases the $m_{\ell\ell}$ differences under two different model interpretations is shown in Figure ???. Overall, with the complications of model dependent decays, inherently rare production, varying mass orderings, and relative scale between sectors, the search for SUSY is an extraordinary challenge. To discover SUSY one should design a search to encompass a large generalized model space and target generic features rather than highly specific corners.

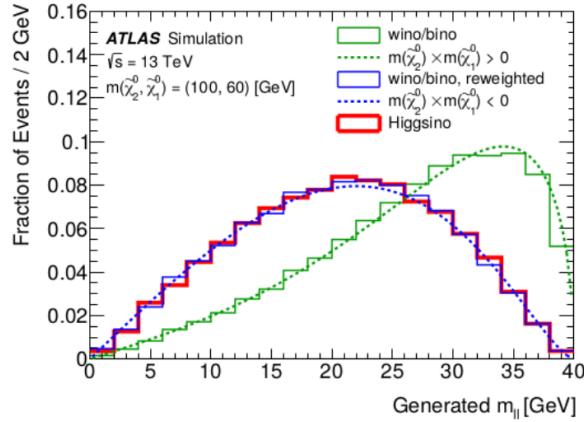


Figure 1.6: m_{ll} reweight with w/b or H interpretations from altas paper in grahams talk

Chapter 2

Motivating the Search for SUSY

2.1 Introduction

The SM is a remarkable theory which describes a wide variety of sub-atomic phenomenon and has consistently held up to tests over many orders of magnitude in energy. However, it's not a perfect theory. There are a few experimental and theoretical problems that the SM can not yet explain like: how to incorporate gravity, how can we explain neutrino mass and mass orderings, why is the universe made up of matter and not antimatter? Observations from the relic microwave background suggest the existence of cold dark matter (cite DM observation), but, there are no suitable SM dark matter candidates to explain the abundance of dark matter. SUSY offers many attractive solutions with the introduction of new particles. One such new particle which can serve as a mediator of gravitational interactions is the gravitino. Other examples include massive invisible particles which act as dark matter candidates, such as the neutralino, χ_1^0 . The neutralino can handle the DM problem with models capable of producing the expected relic DM density of the universe and, in fact, the DM relic density is used to constrain the SUSY model space and simplify searches. Aside from these leading motivations, other more detailed motivations will be discussed in this chapter, the first being the "naturalness problem" with its theoretically aesthetic improvement which adds a symmetry to protect against divergent terms in the perturbative expansion of the Higgs mass. The next motivations are experimental, where SUSY offers an explanation to the significant deviation observed in the muon ($g - 2$) factor from recent FNAL result, as well as

the deviation observed in the W boson mass at CDF. It should be noted that the divergent higgs mass - known as the hierarchy problem - satisfies most SUSY scenarios up to the few TeV scale, but, the two latter experimental measurements motivate searching for SUSY in compressed scenarios.

2.1.1 Stabilizing the Higgs mass

An aesthetic attribute of theoretical models is naturalness, we should expect a model to function "naturally" if the ratio of free parameters in a model are of $O(1)$. Large swings between parameters would be considered fine-tuning and could indicate issues with the underlying theory. So, naturally, if fine-tuning exists in a model, it strongly motivates building extensions to the model to eliminate fine-tuning. One such fine tuning arises in the hierarchy problem, specifically in the Higgs self interaction terms. The SM Higgs Lagrangian terms that involve self interaction are illustrated in equation ??.

$$\mathcal{L} = \frac{gm_h}{4M_W}H^3 - \frac{g^2m_h^2}{32M_W^2}H^4 \quad (2.1)$$

H represents the scalar higgs field, m_h the higgs mass, and m_W the W boson mass. A correction to the higgs mass can be calculated using standard perturbation theory by evaluating the second term of the Higgs Lagrangian. (cite baer)

$$\begin{aligned} \Delta m_h^2 &= \langle H | \frac{g^2m_h^2}{32M_W^2}H^4 | H \rangle = 12\frac{g^2m_h^2}{32M_W^2} \int \frac{d^4k}{(2\pi)^2} \frac{i}{k^2 - m_h^2} \\ &= 12\frac{g^2m_h^2}{32M_W^2} \frac{1}{16\pi^2} \left(\Lambda^2 - m_h^2 \log \frac{\Lambda^2}{m_h^2} + O(\frac{1}{\Lambda^2}) \right) \end{aligned} \quad (2.2)$$

Here the intergal term is the propagator (cite propagator stuff??) for the exchange of a virtual Higgs and is integrated over phase space. The Λ is known as the scale cutoff parameter and should be interpreted as the scale at which the SM breaks down, possibly near the planck scale $O(10^{19})$ GeV. Notice the leading term Λ^2 indicates that the expansion is

quadratically divergent. The divergent mass correction means there needs to be extremely large cancellations, around 20 orders of magnitude, to maintain $\Delta m_h \propto O(m_h)$. This divergent phenomenon can also be observed with fermion masses, but, chiral symmetry protects the fermion mass from divergence by cancelling out high order Λ terms. SUSY offers a similar protection to the Higgs mass by introducing a symmetry with the additional fermionic and bosonic degrees of freedom leading to similar cancellations and a more natural model.

2.1.2 The Muon Anomalous Magnetic Moment

A major experimental motivation for SUSY lies within the measurement of the muon anomalous magnetic moment. Multiple measurements between two labs, Brookhaven National Lab (BNL) and Fermi National Accelerator Lab (FNAL) have shown significant disagreement with the SM. These experiments measure the muon g factor, or specifically, its deviation from two, $(g - 2)_\mu$. The g factor is related to the electromagnetic coupling of charged particles with the photon and largely depends on the tree level lepton-photon coupling, but, gets small quantum corrections from higher order loops. The largest correction being the single photon loop shown in Figure ???. To predict the g factor, an SM calculation is performed with three types of quantum corrections: Quantum Electrodynanic (QED), Electroweak (EW), and Hadronic. Corrections from the Higgs are neglected because the effects are not experimentally observable. The g-factor prediction starts at exactly 2, with QED, and then involves quantum corrections up to $O(10^{-11})$. The prediction is compared with an experimental measurement at a very high level of precision. If the observation were to deviate from the SM prediction, it can indicate new and unaccounted physics interactions with the SM leptons. The current best $a_\mu = \frac{g-2}{2}$ prediction is reported as $a_\mu = a_\mu^{QED} + a_\mu^{EW} + a_\mu^{\text{Hadronic}} = 116591810(43) \times 10^{-11}$. For each of the a_μ components, the QED component enters at the $O(10^{-3})$ and is known to $O(10^{-11})$. the EW component enters the sum at $O(10^{-9})$ and is known to $O(10^{-10})$. Finally the most complicated hadronic component, contributes at $O(10^{-8})$ and is known up to $O(10^{-9})$. The hadronic contributions

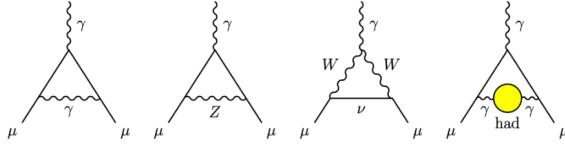


Figure 2.1: figures from <https://www.particlebites.com/?p=8972> which cites pdg

arise from Hadronic vacuum polarization(HVP) and light by light scattering with the HVP diagram also illustrated in Figure ???. The a_μ^{Hadronic} precision dominates the overall a_μ error and is constrained by data driven measurements alongside the limitations of the computational approach with QCD lattice theory. The BNL measurement of a_μ yields a difference with the SM prediction of $\Delta a_\mu := a_\mu^{\text{BNL}} - a_\mu^{\text{SM}} = 279(76) \times 10^{-11}$ which carries significance of 3.7σ . The most recent a_μ measurement from FNAL confirms the BNL measurement within 1σ and the combined experimental average increases the SM deviation with a significance of 4.2σ .

The 4.2σ is a compelling sign for new physics, but not a smoking gun. It is possible to reduce or eliminate the discrepancy by improving the calculations of the HVP and LBL contributions. An early attempt resolve the discrepancy was done by the BMW group and eases the tension to 1.2σ (cite BMW) but still does not fully resolve the differences between observations and theory. If computational improvements can't bring the theory into focus, new particles introduce quantum corrections that will bring experiment and theory into agreement. Several models qualify and successfully explain the a_μ SM deviation, one being SUSY, where for example, contributes additional diagrams via the smuon-muon coupling illustrated in Figure ??.



Figure 2.2: SUSY diagrams explaining g-2 from svens talk

2.1.3 The W boson mass

The W boson is an important and peculiar particle, it is the electrically charged boson and couples only with left handed particles. The decay modes follow two channels: (1) the hadronic mode with different flavor quark pairs and (2) the leptonic mode with a charged lepton and neutrino. Measuring the W mass directly is challenging at the LHC due to either high levels of QCD di-jet background or missing energy from the neutrino. The mass parameter itself, m_W , underpins many important parameters in the SM as well. In fact, m_W is related to the Higgs vev, which implies that coupling of the higg field to all particles is effectively tuned by m_W . Similary, m_W is related to the g factor from $(g - 2)_\ell$ such that $m_W = gv/2$. The W mass can also be paramterized at tree level in terms the fine structure constant α , the Fermi constant G_μ and the Z-boson mass m_Z , with higher order radiative corrections coming from Δr (cite w mass prediction paper)

$$m_W^2 = m_Z^2 \left(\frac{1}{2} + \sqrt{\frac{1}{4} - \frac{\pi\alpha}{\sqrt{2}G_\mu m_Z^2}(1 + \Delta r)} \right) \quad (2.3)$$

There is no exact SM prediction of the W mass, but, since there is an interdependence of many parameters such as v , m_z , G_μ , α , the SM "prediction" is constrained by experimentally measured parameters. The most recent measurement of m_W was performed by CDF II at the Tevatron where m_W was obtained by fitting the kinematic distributions of light leptonic decays recoiling against a system of jets. This measurement is 50% more precise than the previous measurement by ATLAS (cite atlas) and heavier than the SM prediction. The combination of a large deviation with very small error bars results in a significance of 7σ . (cite CDFII).

If we believe the CDF measurement, and follow up experiments confirm the excess in the W mass, it is definite sign of new physics. The new physics would express itself as new particles in the radiative corrections via equation ???. There are numerous SUSY models that could explain the excessive mass of the W boson, (CDF lightsusy) but in general, a

slightly heavier W favors light SUSY models which is illustrated in Figure ???. A light SUSY implies models that are characterized by electro-weak scale SUSY particles among which can favor compressed scenarios. The available parameter space, with an abundance of models, where also CDF II's m_W and the $(g - 2)_\mu$ can be satisfied is shown in Figure ??.

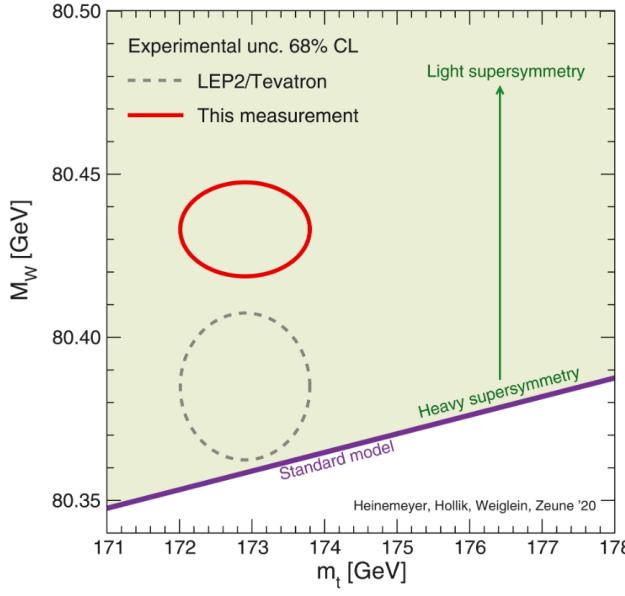


Figure 2.3: plot from CDF paper

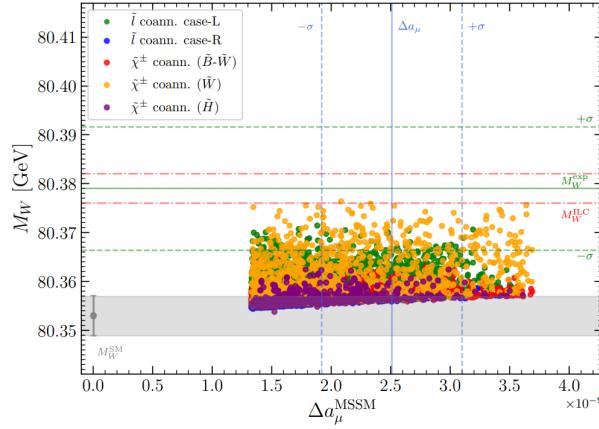


Figure 2.4: svén plot from paper

2.2 The current status of SUSY

There have been many searches for SUSY particles, starting from searches at LEP and still ongoing at the LHC today. There is no observed evidence of SUSY yet, but, there is still not enough lack of observation to fully reject the SUSY hypothesis. The most widely searched region SUSY space is related to strong production of SUSY particles. The large expected cross sections compared to other sectors gauginos/sleptons offers the most low hanging fruit for potential discovery. Simplified models in CMS have excluded \tilde{g} and \tilde{q} up to around 2 TeV with the most recent limits are shown in Figure ???. The area inside the lines in Figure ?? indicate that the 2-D mass points of the sparticle and LSP pair are ruled at a 95% confidence level. Similarly the CMS slepton and electroweak limits are shown in Figure ???. Note that

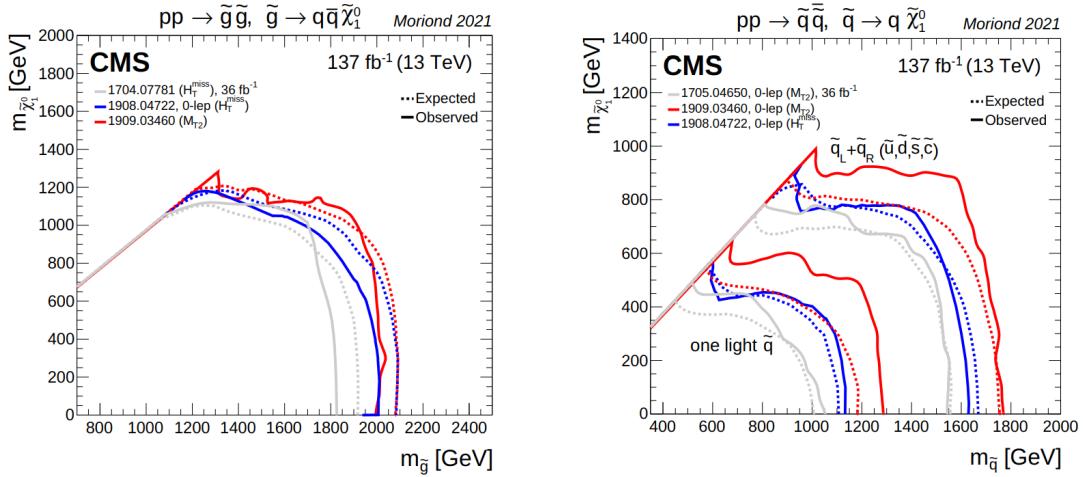


Figure 2.5: strong limits

the electro weak limits are just now reaching the TeV scale while SUSY remains valid at the few TeV level, so, there is plenty of room in the 2-D mass plane to either discover or exclude SUSY by adding more data. Notice also that the corridors with compressed regions for electroweakinos and sleptons are effectively untouched. So, a well motivated region to search is near the LSP-sparticle mass degeneracy line in the 2-D mass plane because it is both unexplored at high masses and theoretically and experimentally well motivated.

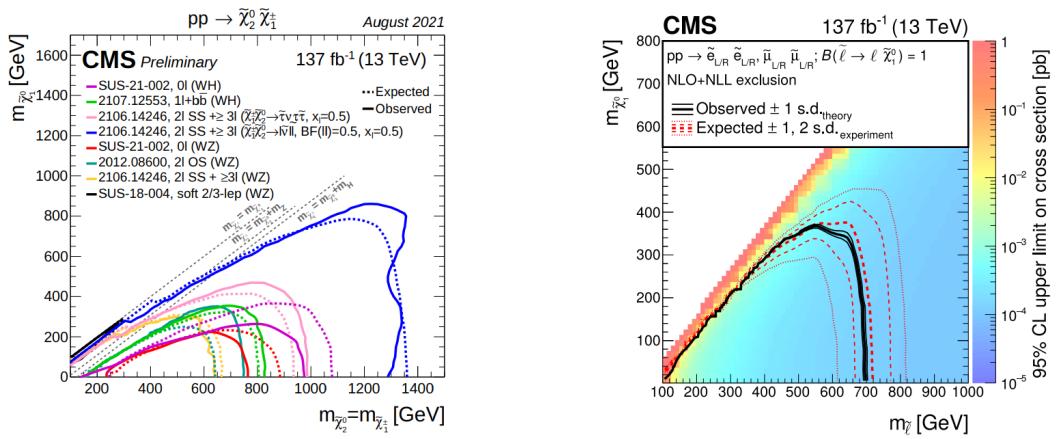


Figure 2.6: limits

Chapter 3

The CMS experiment

3.1 Introduction

The Compact Muon Solenoid (CMS) experiment consists of a detector which is a part of Large Hadron Collider (LHC). The detector encapsulates two synchronous bunches of high energy protons which counter rotate through the LHC accelerator ring. The protons collide at the center of the detector with a significantly large energy and the expectation that more massive and potentially new particles will be produced. Each particle produced in the collision can either decay, interact, or escape detection. The particles that interact are measured by the detector, where different layers specialize in measuring certain classes of particles. Then, from final state energy and momentum measurements, the initial interaction and everything in between is reconstructed.

3.2 The Large Hadron Collider

The LHC is a circular collider designed to collide proton beams with a centre-of-mass energy of 14 TeV and an instantaneous luminosity of $10^{34} \text{ cm}^{-2}\text{s}^{-1}$. (cite lhc paper direct quote). The main accelerator ring consists of two counter rotating proton beams which are incased in an ultra high vacuum to prevent unintended interactions. The beams are accelerated with cryogenic electro-magnets which operate at -273°C and are cooled by liquid helium. There are two types of magnets present, 1232 dipole magnets which bend the beam around the ring and 392 quadrupole magnets which focus the beams. The beam itself is structured with

proton bunches, with each bunch spaced 25 ns apart and 2808 bunches per beam. The period of recording collision data are referred to as runs, of which, there are two completed runs, denoted as Run I, and Run II with integrated luminosities 58 fb^{-1} and 138 fb^{-1} respectively. There is also an expected cumulative integrated luminosity of up to 500 fb^{-1} which includes Run I, Run II, and the presently ongoing Run III.

3.3 The CMS Detector

The CMS detector is a hermetic shell that encapsulates the two counter rotating proton beams. The beams collide at the center of the detector and produce outgoing showers of particles that travel transverse to the beam axis. The observable outgoing particles, depending on the type of particle, are then measured in one of the specialized concentric layers of the detector. The initial transverse depiction of sub atomic interaction, and intermediate particles, can then be reconstructed from the energy and momentum measured in the detector. The total longitudinal momentum is not reconstructable for two reasons: first being that the momentum fraction of the initial partons is unknown and second is that some particles travel along the beam line outside detector acceptance. There are an abundance of collisions seen by the detector but not every event is recorded. Instead, interesting events, say due to the presence of a muon or large missing energy, trigger the detector to take a snapshot and permanently record said interesting event.

The chronology of a particle traversing the detector, as shown in Figure ??, is as follows. Particles are produced post-collision at a primary interaction point, or primary vertex. Other interactions can occur nearby in the same snapshot and are denoted as pile-up, which is a form of noise obfuscating the primary interaction. From either primary or secondary vertices, both charged and neutral particles traverse the first region of the detector, the silicon tracker. The silicon tracker consists of concentric thin electronic sensors that register "hits" from only charged particles. Each sequence of hits can be connected into a "track" that represents the

path and origin of the charged particle. The next stop for particles, is the Electromagnetic Calorimeter (ECAL). The ECAL consists of scintillating PbWO₄ crystals that are designed to stop and measure the energy deposits of photon and electrons. The energy deposits from the two are distinguished by tracks that seed ECAL showers. Anything that makes it through the ECAL, encounters the hadronic calorimeter (HCAL). The HCAL consists of brass and plastic scintillators that stop the remaining massive particles and measures their energy. The last two regions of the detector are generally only seen by muons and are the centerpieces of CMS. First is the the solenoidal magnet, which generates a 4 Tesla uniform magnetic field throughout all of the inner regions of the detector. The magnetic field allows the measurements of two important observables: charge and momentum. A charged particle's path will bend in the presence of a magnetic field, and the clockwise or counter clockwise trajectory indicates the charge, while the curvature of the bend determines the momentum. The outer-most part of the detector is the muon chamber, which similar to the tracker, registers a sequence of hits via drift tubes or cathode strips. The tracks in both the tracker and muon chambers can then be combined to precisely measure the momentum of the muon in addition to resistive plate chambers which act as a hardware level muon trigger.

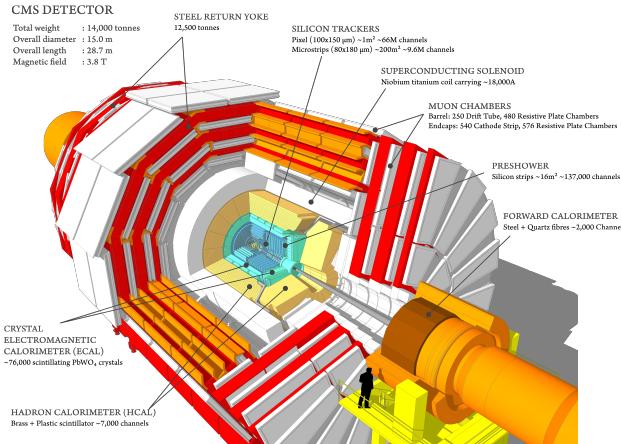


Figure 3.1: plot from CMS of CMS

Chapter 4

Compressed SUSY Search

4.1 Compressed Search Introduction

In accordance with the strong experimental and phenomenological motivation for compressed spectra accompanied by the pursuit to comprehensively test SUSY, we conduct a search designed to be generic sensitive to many final states that involve missing energy. The targeted processes include, but are not limited to the production of stops, electroweakinos, and sleptons with diagrams included in Figure ???. The common thread between all of these

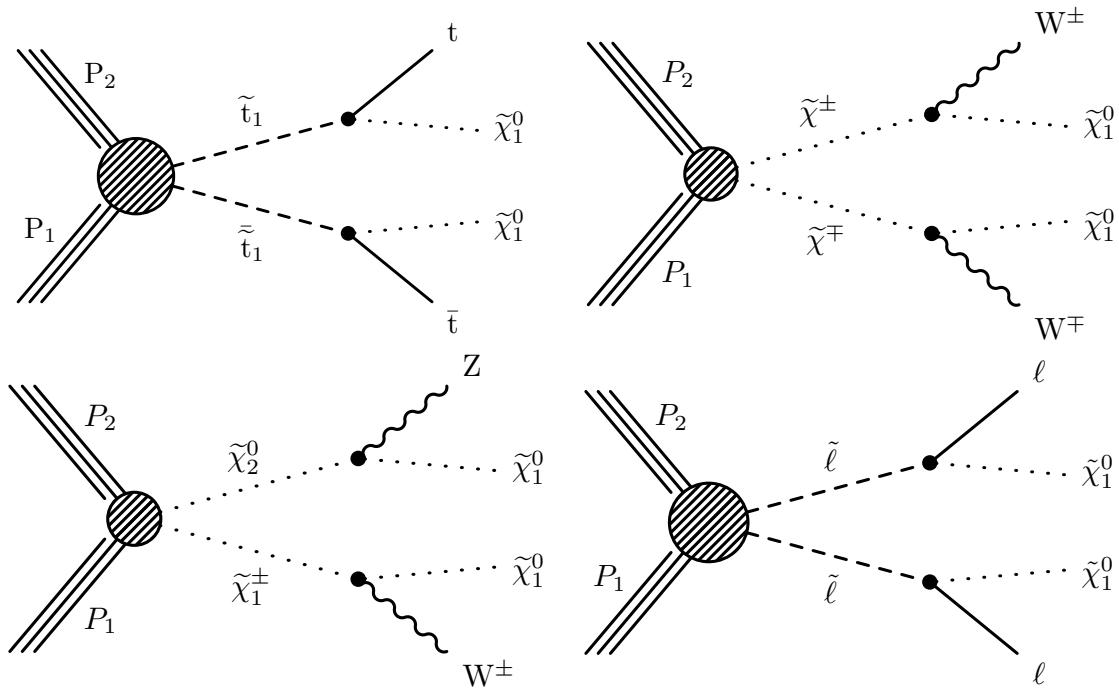


Figure 4.1: diagrams

processes is a pair produced visible system alongside a massive invisible system. In the case

of a compressed scenario, most of the energy available in the system is used by the rest mass of the LSP. These small mass splittings leads to low momentum visible products that are difficult to reconstruct or are undetectable. In the case of intermediate massive particles, such as W or Z boson, these are forced off-shell so the visible products receive even less momentum, as the available energy goes into the mass of the intermediate particle. In order to identify these type of events we study cases with significant initial-state radiation (ISR). The ISR system recoils against, or boosts, the sparticle system, leading to high missing transverse momentum which is a tractable experimental signature. A depiction of this type of event is shown in Figure ???. With the ISR assisted pair produced topology we categorize and subdivide the visible and invisible system using Recursive Jigsaw Reconstruction (RJR) to approximate various rest frames. From those rest frames, we compute a basis of kinematic observables that describe features aid in the discrimination of compressed SUSY against SM processes.

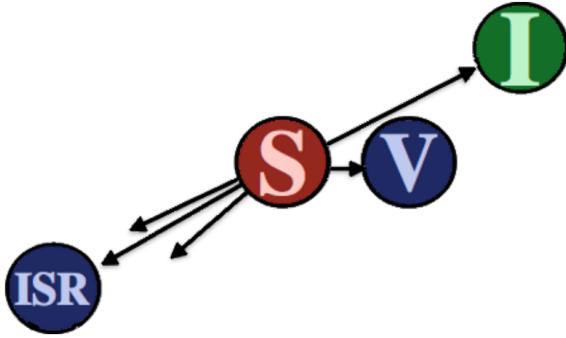


Figure 4.2: recoiling system

4.2 RJR Reconstruction

The ISR assisted topology involves a collimated invisible and soft visible system together recoiling against another visible system of ISR jets. Each event is organized by imposing a decay tree onto the visible (V) and invisible objects (I) and assigning the visible components to either the ISR or sparticle (S) side of the event. From the initial assignment, the sparticle

system is further subdivided into subsystems A and B where each sub system has a visible ($V_{a/b}$) and invisible ($I_{a/b}$) component. The three decay trees that contains these sets of objects in different reference frames are illustrated in Figure ???. In order to resolve the two invisible

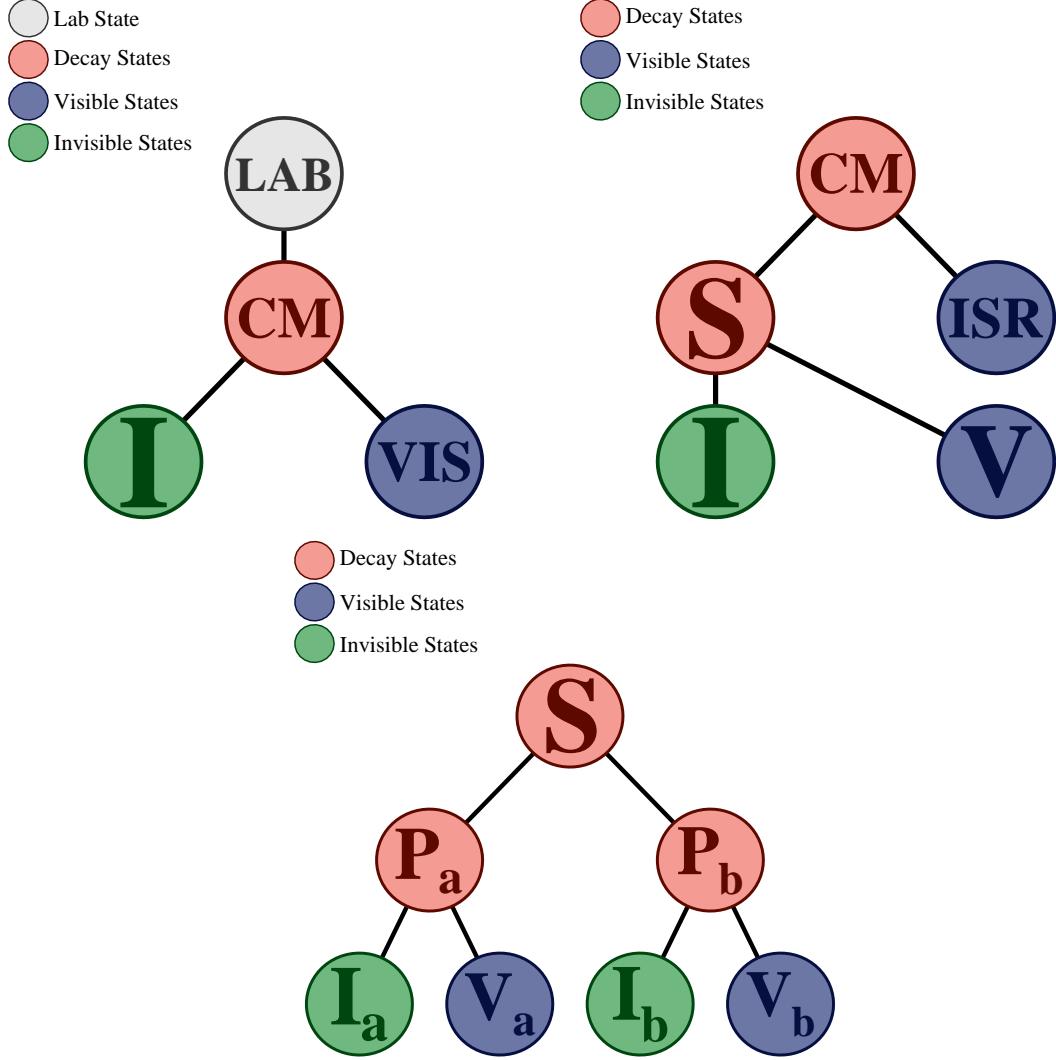


Figure 4.3: decay trees

four momenta, kinematic and combinatoric unknowns need to be estimated. For example, the visible products are indistinguishable, so, there needs to be a metric which dictates the assignment of the visible objects to either the ISR or sparticle system. Similarly, the sparticle subsystem partitioning, both visible and invisible, must be determined when only the transverse invisible momentum is known. Thus, the combinatoric assignment and the

estimated four momenta depend on each other, so, we apply a set of rules to simultaneously determine both. RJR provides the framework to organize and evaluate each event. The set of rules used by this analysis are as follows:

1. Assign charged leptons to the S system
 - Target leptonic sparticle signatures and fixing the leptonic system to always recoil against ISR
2. Fully determine the set of $\{V, ISR\}$ objects by assigning other visible objects to either the S or ISR systems by maximizing the momentum of the sparticle system in the CM frame

$$\{V, ISR\} = \arg \max_{V, ISR} p_S^{CM} \quad (4.1)$$

3. The visible S system objects are assigned to V_a or V_b by minimizing the mass of the sparticle subsystems i.e. grouping objects that traveling in similar directions

$$\{V_a, V_b\} = \arg \min_{V_a, V_b} M_{P_a}^2 + M_{P_b}^2 \quad (4.2)$$

4. Adjust the total mass of the invisible system according the visible systems where the individual invisible masses are constrained to zero.

$$M_I^2 = M_V^2 - 4M_{V_a}M_{V_b} \quad (4.3)$$

5. Estimate the longitudinal component of invisible momentum by minimizing CM mass, i.e. determining the transverse mass of V+I systems

$$\vec{\beta}_{CM,z}^{\text{lab}} = \arg \min_{\vec{\beta}_{CM,z}^{\text{lab}}} M_{CM} \quad (4.4)$$

6. Determine the full kinematics of I_A and I_B and momentum partitioning by evaluating

the S frame velocities also through minimizing the mass of the sparticle subsystems

$$\vec{\beta}_{P_a}^S, \vec{\beta}_{P_b}^S = \arg \min_{\vec{\beta}_{P_a}^S, \vec{\beta}_{P_b}^S} M_{P_a}^2 + M_{P_b}^2 \quad (4.5)$$

By recursively iterating through the various combinations of objects and determining the four momenta of the groupings from Figure ??, the optimal organization for the event is determined that satisfies the aforementioned RJR prescription. The consequences of this organization is discussed in the following section where we construct a basis of kinematic variables to exploit characteristics of compressed SUSY and discriminate against SM processes.

4.3 Compressed Kinematics

The main observables are designed to be sensitive to the properties of compressed SUSY. One of these properties is the mass of the invisible particle in the event. With the ISR assisted system recoiling against a massive invisible particle, the invisible gets a large momentum kick from ISR. This characteristic can be exploited with the variable R_{ISR} which is defined as:

$$R_{ISR} = \frac{|\bar{p}_I^{CM} \cdot \hat{p}_{ISR}^{CM}|}{|\bar{p}_{ISR}^{CM}|} \sim \frac{m_I}{m_P} \quad (4.6)$$

In a compressed scenario, the fraction of the invisible momentum to the total ISR kick would be expected to be close to one, as the visible system uses a small fraction of the momentum of the event and the ISR vs invisible system would be nearly anti parallel. The peak of the R_{ISR} distribution can be approximated by the ratio of the true invisible mass to its sparticle parent mass. This behavior of R_{ISR} is that is spread in the distribution is reduced and the peak approaches one as the level of compression increases. Similarly the R_{ISR} behavior doesn't depend specifically on the underlying SUSY process, but only that there is a heavy invisible system recoiling against ISR. SM backgrounds do not exhibit the same behavior in

R_{ISR} , so, the result is strong discrimination between compressed SUSY and SM at high R_{ISR} which actually improves as the mass splittings $\Delta m = m_P - m_I$ decrease. An illustration of the R_{ISR} shapes comparing signal to background is show in Figure ??

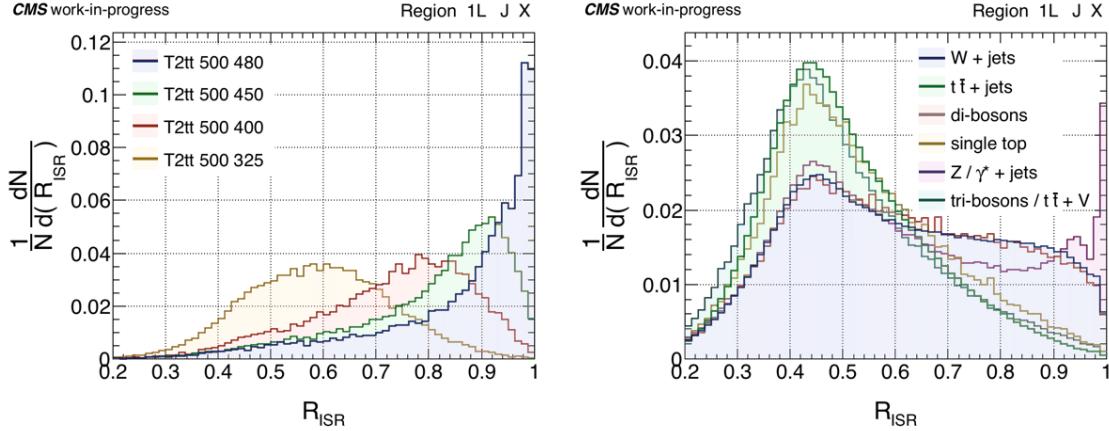


Figure 4.4: Risr distributions 1L selection di stop

Another observable which is uncorrelated with ISR but also sensitive to compressed topologies is M_\perp . M_\perp is constructed from the average squared masses of the sparticle subsystems $M_{P_a/b}$ and is explicitly defined as

$$M_\perp = \sqrt{\frac{M_{P_a\perp}^2 + M_{P_b\perp}^2}{2}} \quad (4.7)$$

The individual invisible masses are constrained to zero, and, in the case of massive invisible particles is an incorrect assumption. The consequence is then that M_\perp is sensitive to the inherent mass splittings between the parent sparticle and LSP. The behavior of the M_\perp distribution is that it exhibits a kinematic endpoint or edge at the $\Delta m = m_P - m_I$ and gives the strongest discrimination against SM processes at larger values of M_\perp , which may still be compressed e.g. in case of intermediate top quarks. An example of the shape of the M_\perp distributions in with signal versus background is shown in Figure ??

The combination of both R_{ISR} and M_\perp form a 2-D plane in which to conduct a "bump hunt" which is illustrated in Figure ???. Here the localization of the bump in the plane depends on

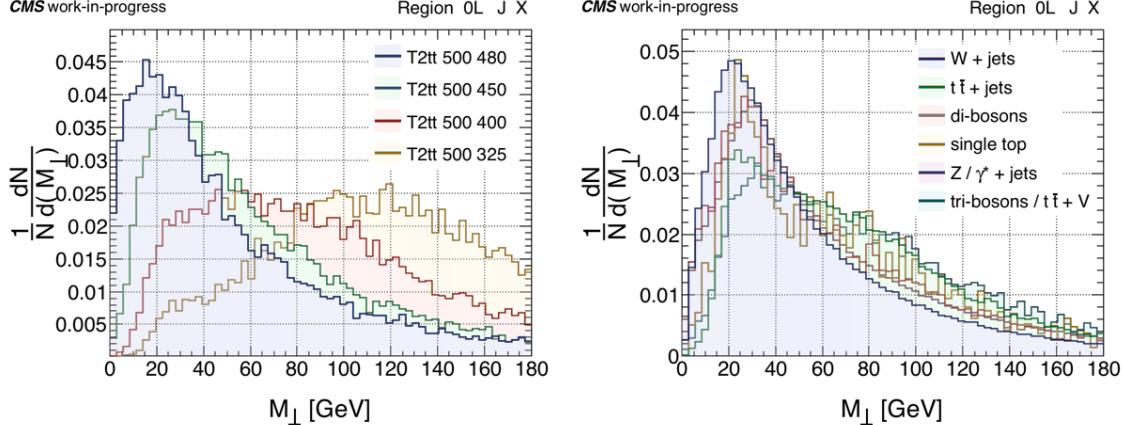


Figure 4.5: mperp shapes

the sparticle masses, the larger both masses the less over spread in the distribution. Also the smaller the mass splitting between sparticles, the peak of the distribution is pushed sharply towards 1 in R_{ISR} and the stronger the discrimination against SM backgrounds becomes. Binning in this 2D plane gives us the most sensitive signal region at high R_{ISR} while the opposite, low R_{ISR} provides a background rich region to constrain the background yields in the sensitive region.

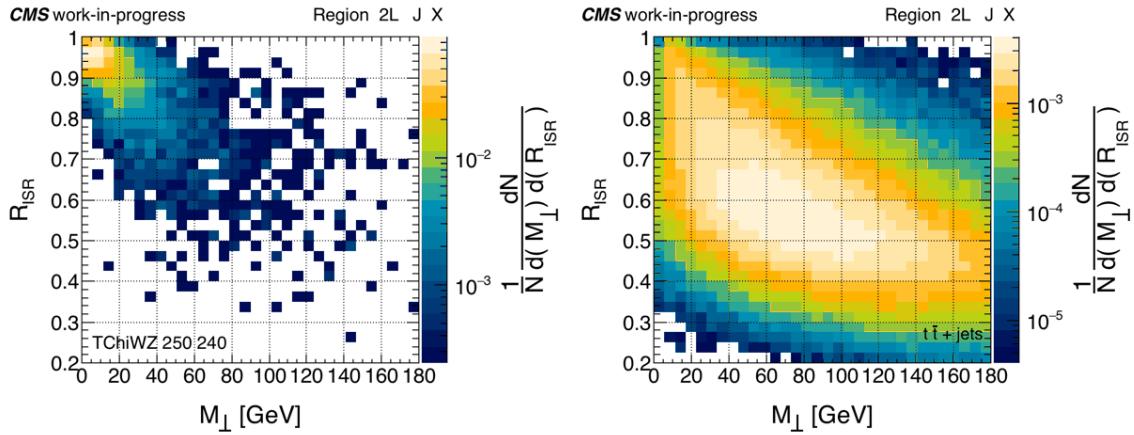


Figure 4.6: 2d plots

Two additional kinematic variables are utilized to aid in discrimination against SM backgrounds, both are less powerful than R_{ISR} and M_\perp but still very useful. The first quantity is complementary to R_{ISR} and is the transverse momentum of the ISR system p_T^{ISR} . The

more momentum in the ISR system means the sparticle system gets kicked harder - leading to better resolution in the R_{ISR} distribution from the p_T^{ISR} R_{ISR} correlation. Fortunately the R_{ISR} distributions from backgrounds are anti correlated with p_T^{ISR} which means that the combination of both high p_T^{ISR} R_{ISR} provides a rich signal region. Both correlations for signal and background can be visualized in Figure ??

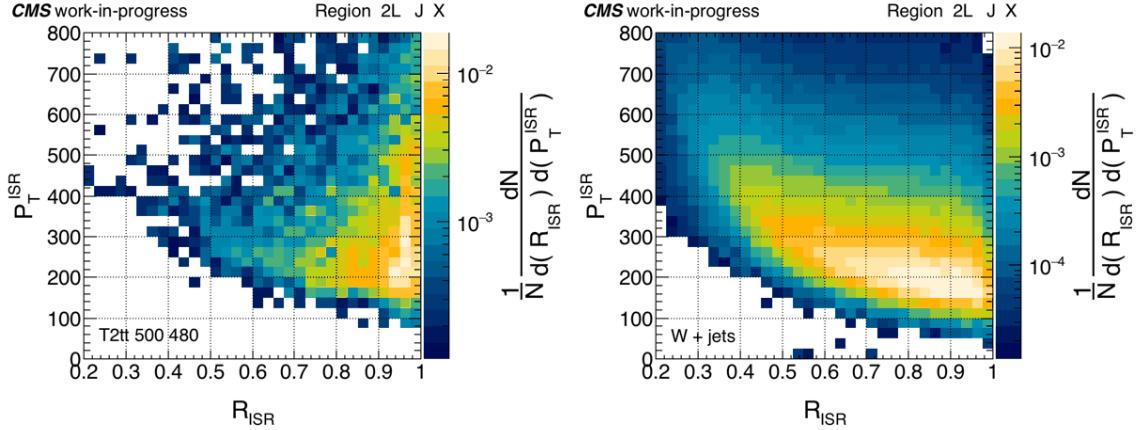


Figure 4.7: pt_{ISR} $risr$ distributions

The other kinematic variable is γ_\perp which is a measure of the symmetry of the di-sparticle system and is defined as:

$$\gamma_\perp = \frac{2M_\perp}{M_{S\perp}} \quad (4.8)$$

Here M_S is the mass composed of the transverse four momenta of all objects from both sparticle subsystems. The behavior of the mass ratio γ_\perp is that it tends to larger values asymmetry in the final state, as illustrated in Figure ???. This is useful because it better isolates signals and backgrounds with pairs of W or Z bosons. Since both complementary variables p_T^{ISR} and γ_\perp don't have the discriminating power of R_{ISR} and M_\perp we categorize events. With p_T^{ISR} we have high or low categories, where the lower edge and low to high pivot depends on the lepton and jet multiplicity of the event. In the case of γ_\perp we also construct high and low categories about the value $\gamma_\perp = 0.5$. The combination of high and low and the divides selected for each pair of categories is designed to create more signal sensitive high regions with background yields that are constrained by low background rich categories.

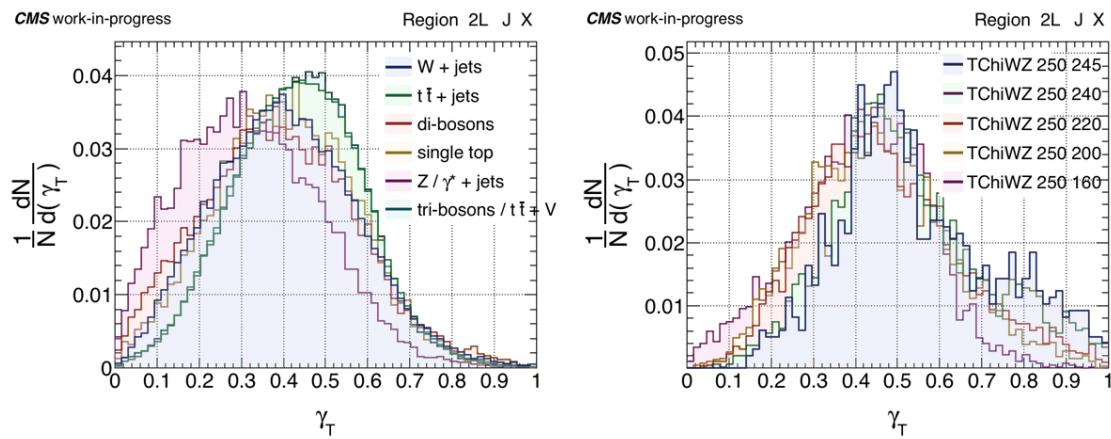


Figure 4.8: gamT dist

Chapter 5

Analysis Description

5.1 Introduction and Strategy

The full analysis is built on the compressed kinematics described in the previous chapter. This goal of this chapter is to outline specific details and the strategies used to potentially discover SUSY. This includes the description of events selected to analyze and the objects that compose these events. This search casts a wide net to capture a wide variety of signatures and final states, the consequence of this strategy yields a large number of categories and bins that cover many multiplicites of lepton and jet final states. Finally, I will discuss the data driven approach to constrain and predict background events in the most sensitive regions by conducting a series of fits to construct a robust fit model.

5.2 Data and Simulation

The analysis involves the full Run II dataset which is divided into three subsets by the years 2016,2017,2018 and total integrated luminosity of 138fb^{-1} . Each year is comprised of $36.31\text{ fb}^{-1} \pm 1.2\%$ (cite lumi 17 003), $41.48\text{ fb}^{-1} \pm 2.3\%$ (cite lumi 17 004), $59.83\text{ fb}^{-1} \pm 2.5\%$ (cite lumi 18-002) in 2016, 2017, and 2018 respectively. The data is modeled by MC that represents the full SM background and is qualitatively grouped by process and final state. These grouping of SM backgrounds is defined in table ?? along with the partnered signal models.

The signals that will be addressed in this work include mutiple different sparticles and final

Table 5.1: table caption

Bkg. Label	Bkg. Composition
W + jets	Single W boson, a dominant background that composes about 50% of the total background
tt+jets	$t\bar{t}$ which can be accompanied by a W,Z,h, or γ , the other dominant background composes about 50% of the total background
ZDY Di-boson (DB)	Z+jets and Drell Yan, an intermediate background WW,ZZ,WZ,Wh,Zh, an intermediate background
ST	Single top processes including tW, a rare background
Tri-boson (TB)	WWW,ZZZ,WWZ, WZZ, WZ γ , WW γ , a rare background
Signal Label	Signal Composition
T2tt TChiWZ TSlepSlep TChipmWW	$pp \rightarrow \tilde{t}\tilde{t}; \tilde{t} \rightarrow t\tilde{\chi}_1^0$ $pp \rightarrow \tilde{\chi}_2^0\tilde{\chi}_1^\pm; \tilde{\chi}_2^0 \rightarrow Z\tilde{\chi}_1^0; \tilde{\chi}_1^\pm \rightarrow W^\pm\tilde{\chi}_1^0$ $pp \rightarrow \tilde{\ell}\tilde{\ell}; \tilde{\ell} \rightarrow \ell\tilde{\chi}_1^0$ $pp \rightarrow \tilde{\chi}_1^\pm\tilde{\chi}_1^\mp; \tilde{\chi}_1^\pm \rightarrow W^\pm\tilde{\chi}_1^0$

states, which also will include multiple interpretation of their analysis results. A list of the signals is provided in table y. Each signal is produced according to an (LSP,NLSP) mass grid for each year. The raw number of events per mass points and grid spacings for the signals shown in table ?? are displayed in Figure 1.1 with all years combined.

The majority of signal and backgrounds use the MadGraph [?] generator to model at LO and NLO. The ST backgrounds use PowHEG 2.0 [?] to model at NLO. Parton shower and fragmentation for all samples is done with PYTHIA 7 [?]. Each year is subjected to an underlying event tunes with CUETP8M1 for 2016, CP2 for 2017 and 2018 signals and CP5, for 2017 and 2018 backgrounds [?]. The detector conditions and response are simulated for all samples with GEANT4 [?].

5.3 Event Selection and Physics Objects

For events to be qualified for analysis, they pass a handful of selected triggers and preselection which reflect the compressed kinematic description provided in the previous chapter. The

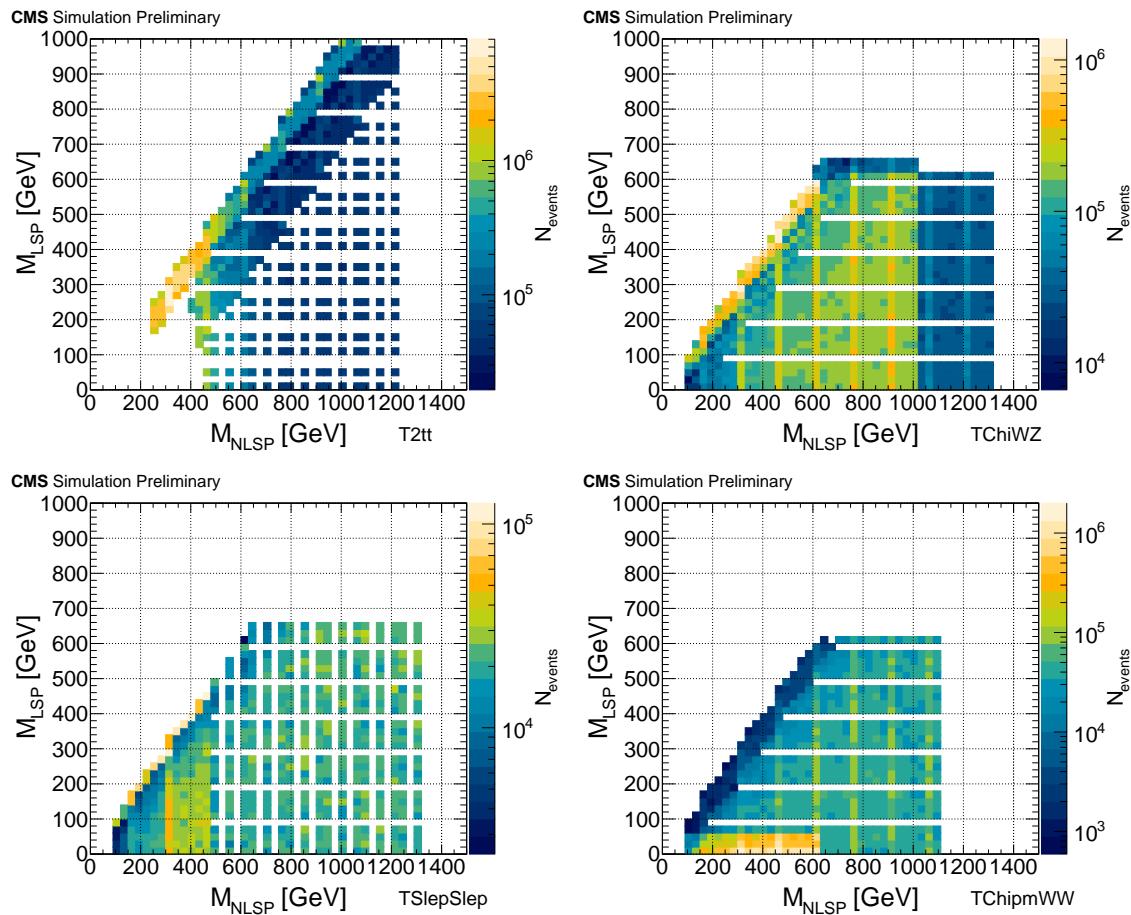


Figure 5.1: grids

triggers used PFMET PFMHT cross triggers. PFMET is the particle flow missing transverse energy which is expected to capture the p_T^{Miss} from the LSP. PFMHT is expected to trigger on events with significant jet activity and multiplicity which are likely candidates for ISR events. In general, the compressed SUSY topology is somewhat rare organization of an event, due to the uncommon nature of these types of events the MC modeling does not always sufficiently and precisely describe data, so, a data driven approach is utilized for physics objects to compensate for any disagreement. This approach compares the overall efficiency or behavior of selected objects and computes data driven scale factors, while also providing a platform to model and understand systematic effects. These calibrations are then applied to MC to bring data and MC into agreement. The instance of Scale factor generation arises in the modeling of the efficiencies of the trigger turn ons, with efficiency defined as events that pass the trigger and preselection versus only preselection. The comparisons of data with which is shown in Figure 1.2 with the efficiency shapes modeled by a Gaussian CDF.

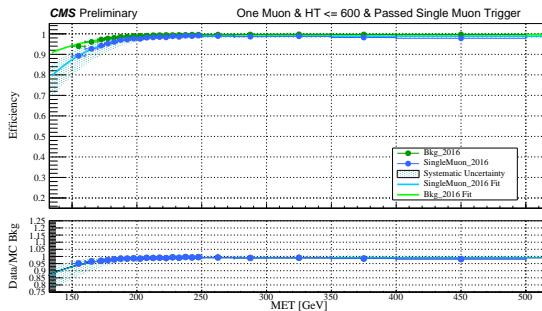


Figure 5.2: met trig sf

The aforementioned pre selection which is a set of carefully studied criteria to remove background and mismodeled events as well as select events with objects that are expected to populated targeted final states. The pre selection consists of the criteria listed in the following Table Z:

Preselection Requirements

Criteria	Description
$N_V \geq 1$	At least one visible object assigned to the S system
$N_j^{ISR} \geq 1$	At least one jet assigned to the ISR system
$p_T^{miss} > 150\text{GeV}$	Minimum transverse missing energy based on trigger efficiency
$p_T^{ISR} > 250\text{GeV}$	Minimum ISR kick to resolve massive invisible particles
$R_{ISR} > 0.5$	Target Massive LSPs
$ \Delta\phi_{\vec{p}_T^{miss}, V} < \pi/2$	Ensures visible and invisible system are traveling in the same direction
$p_T^{CM} < 200$	Rejects mismodeled events
veto $f(\Delta\phi_{CM,I}, p_T^{CM})$	2D function to also reject mismodeled events - See Fig 1.3

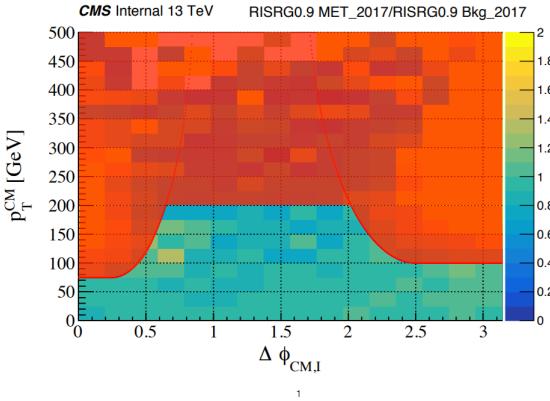


Figure 5.3: cleaning cuts

The pre selection forms the basis for an event to be analyzed. Following preselection, the physics objects can be selected, classified, and categorized. The possible object composition can consist of jets, b-tagged jets, soft secondary vertices (SVs), and leptons. The discussion of the leptons and their classification will be reserved for the following chapter alongside the calculation of lepton scale factors. A summary of these physics objects and their kinematic requirements are listed in the following Table X. The various types of jets involve working points (WP) from their respective physics object groups and are standard objects used in

Table 5.2: table caption2

Visible Physics Objects	
Jets	AK4 PF Jets Tight ID $p_T^{jet} > 20$ GeV $ \eta > 2.4$
B-tagged Jets	AK4 PF Jets DeepJet Medium WP $p_T^{b-jet} > 20$ GeV
SVs	$2 < p_T^{SV} < 20$ GeV
Leptons	Very Loose ID $p_T^{\mu^\pm} > 3$ GeV $p_T^{e^\pm} > 5$ GeV Gold/Silver/Bronze quality classes

CMS physics analysis (cite jet stuff ak4 and ids). The b-tagging is done by a standard NN based tagger which only identifies b-jets down to 20 GeV. A complementary SV tagger was developed specifically for this analysis and a detailed description of this tool is described in this analysis's sister thesis (cite erich thesis). The purpose of the SV tagger is to efficiently extend the b-tagging range down to 2 GeV because a final state topology with something like compressed stops often includes soft b-jets.

5.4 Categorization and Fit Strategy

Once an event passes all the preselection and the objects are classified, the event is then categorized based on object composition, object multiplicity, and kinematic characteristics.

Chapter 6

The Tag-and-Probe

Abstract

The Tag-and-Probe is a method used to measure the selection efficiencies of an object using data. In the context of this compressed SUSY analysis, the Tag-and-probe measures the efficiencies separately of each light lepton(e/μ) selection criteria. The total lepton selection efficiency is then computed by combining factorized efficiency components. The same general method is used for both electrons and muons, however, Muons utilize the J/ψ di-muon trigger which allow more precise efficiency measurements from data at lower p_T .

6.1 Introduction and Methodology

An important element of a lepton based search is properly modeling the efficiency of selected leptons. A purely Monte-Carlo driven approach is inadequate in perfectly describing nuances in data due to imperfections in modeling. Instead of trying to model exactly all physics and detector effects with simulation, the efficiencies can be directly measured from data by using the Tag-and-Probe method.

The Tag-and-Probe method is used to measure a selection criteria by using a well known resonance such as a Z , J/ψ , or Υ and counting the number of probes that pass that criteria. Each counted instance of the Tag-and-Probe consists of two selected leptons. One of the selected leptons is the tag and the other is the probe. The tag passes tight selection require-

ment to give high confidence that it isn't a fake lepton. Fake leptons fall into two possible categories: reducible and irreducible. A reducible fake lepton is a particle that fakes the signature of a lepton such as a charged pion. An irreducible fake lepton is an actual lepton which coincidentally passes some selection criteria but is not the targeted leptons of interest e.g. an isolated muon from a jet accompanying a leptonic Z decay of interest. The second lepton in the Tag-and-Probe is the probe. The probe is subjected to the selection criteria whose efficiency is being measured. The invariant mass of the pair of leptons is calculated and required to fall within a defined range around the resonance. A particular event may have multiple lepton pairs but the tag and the probe are not allowed to switch positions and be counted twice, as double counting would lead to a bias in the efficiency measurement [?]. To avoid bias, the tag and probe are required to be the opposite charge and same flavor where the tag is randomly selected. If multiple same flavor lepton pairs occur in single event i.e. there are multiple probes to a single tag, the treatment for selecting the pairs differs between electrons and muons. There is no specific study which led to justifying the differing arbitration approaches in flavors, only that the choice reflects the default choices implemented in the existing code bases. For muons, no arbitration is used, all pairs are utilized which means an additional pair not truly from the resonance will then contribute as combinatorial background in a single event. For electrons, only a single probe is selected per event which has the highest p_T . The selected probes can either pass or fail their selection which leads to the formation of three distributions, one with a passing probe, one with a failing probe, and one with all probes. An example of all three distributions is shown in Figure ???. The probability of observing k passing probes in n Tag-and-Probe pair trials is dependent on the selection efficiency ε and can be expressed as a likelihood from the binomial probability density $P(k|\varepsilon, n) = \binom{n}{k} \varepsilon^k (1 - \varepsilon)^{n-k}$. The MLE estimator for efficiency is then the fraction of passing probes to the total number of pairs, or $\varepsilon = k/n$. Technical documentation for the Tag-and-Probe in CMS is scarce, but, an early strategy for fitting efficiency is defined in [?]. The legacy code base as of `CMSSW_10_6_X` uses a binned maximum likelihood between the

observed passing probes and failing probes where the efficiency extracted is an explicit fit parameter. The two simultaneously fit functions are:

$$N^{\text{Pass}} = N_{\text{Total}}(\varepsilon \cdot f_{\text{All}}^{\text{sig}}) + \varepsilon_{\text{bkg}} \cdot (1 - f_{\text{All}}^{\text{sig}})) \quad (6.1)$$

$$N^{\text{Fail}} = N_{\text{Total}}((1 - \varepsilon) \cdot f_{\text{All}}^{\text{sig}} + (1 - \varepsilon_{\text{bkg}}) \cdot (1 - f_{\text{All}}^{\text{sig}})) \quad (6.2)$$

$N^{\text{Pass/Fail}}$ is the total number of observed probes that either pass or fail the selection criteria while N_{Total} is the total number of Tag-and-Probe pairs. The binomial estimator for efficiency, ε , enters the fit functions as the first term but is accompanied by a second term that describes the background contribution with its own efficiency ε_{bkg} . The term $f_{\text{All}}^{\text{sig}}$ is the fraction of background subtracted signal events over the allowed dilepton mass range. $f_{\text{All}}^{\text{sig}}$ depends on the defined signal and background pdfs. The nominal pdfs chosen for reported fits uses a 5 parameter Voigtian+Voigtian signal model which share a common mean but use independent Γ and σ . The signal model is combined with an Exponential background model.

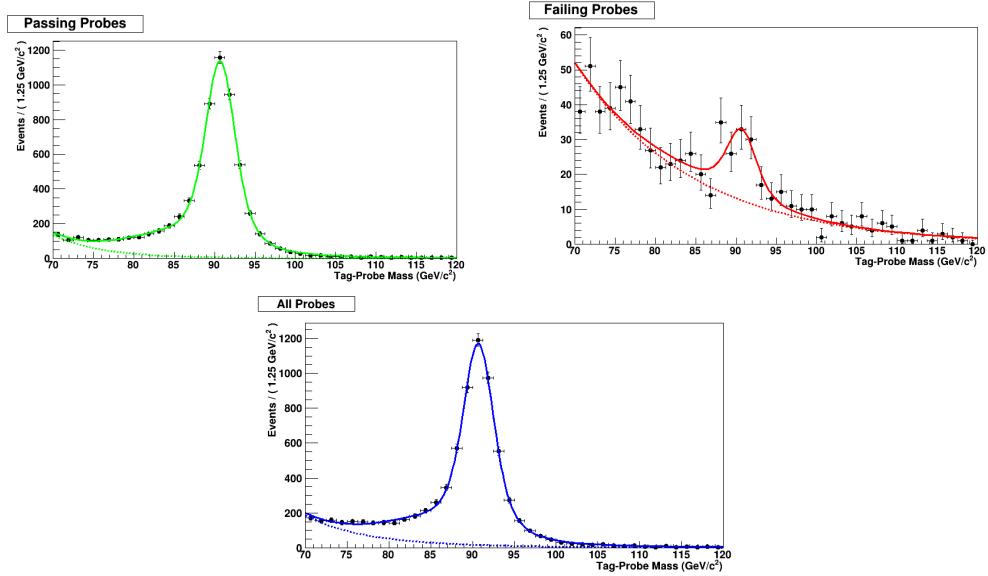


Figure 6.1: Example Tag-and-Probe Z di-muon fits for passing,failing, and all probes with the Medium Id, $|\eta| < 1.2$, and $p_T < 20$ GeV

6.2 Lepton Object Definitions

Leptons are selected according to the minimum requirement “VeryLoose” which depend kinematic and topological quantities which are shown in Table ???. The electrons use an additional loose MVA requirement: MVA VLooseFO ID [?]. The set of VeryLoose leptons are further subdivided by quality into three mutually exclusive categories: Gold, Silver, and Bronze. Each category has a measure of three main quantities, the first being the quality of the pre-determined Id. The Id’s differ per flavor and are the standard working points defined by the corresponding physics object group. The muons use the Medium Id [?] and electrons use a more strict selection, due to their messy nature, with the Tight Id [?]. The second quantity is the “promptness” or distance of the lepton production point from the primary vertex. Promptness is measured by the significance of the 3D impact parameter (SIP3D) which is defined as the impact parameter normalized by its measured error. A $\text{SIP3D} > 1$ is associated with a secondary particle which is not produced at the primary vertex. The last component is the isolation, a measure of the density of particles in a cone around the lepton. Two similar but complimentary absolute isolations are used: PFIso [?] and MiniIso [?]. Both isolations are an energy sum of neighboring particles inside a cone, but, PFIso has a fixed cone size of $R = 0.4$ cm and miniIso cone sizes varies inversely with lepton p_T as shown in ??.

$$R_{\text{miniIso}} = \begin{cases} 0.2 & p_T^\ell < 50\text{GeV} \\ \frac{10}{p_T^\ell} & 50\text{GeV} \leq p_T^\ell \leq 200\text{GeV} \\ 0.05 & p_T^\ell > 200\text{GeV} \end{cases} \quad (6.3)$$

Mini isolation also includes effective area pile-up corrections provided in a look up table of bins of p_T and η in the CMSSW Producer/Ntuplizing stage. The implementation of mini-isolation and their corrections utilize the same IsoValueMap producer as used in NANO AOD as of `CMSSW_10_6_X`.

The explicit flavor independent formulas for Gold, Silver, and Bronze can be generalized

by the product of three components which are the measured efficiencies of the three previously mentioned quantities. The efficiencies take the form of conditional probabilities to be measured independently in sequence relative to each other:

$$\begin{aligned}\epsilon_{\text{Gold}} &= \epsilon_{\text{ID}} \times \epsilon_{\text{Isolated}|\text{ID}} \times \epsilon_{\text{Prompt}|(\text{ID} \cap \text{Isolated})} \\ \epsilon_{\text{Silver}} &= \epsilon_{\text{ID}} \times \epsilon_{\text{Isolated}|\text{ID}} \times (1 - \epsilon_{\text{Prompt}|(\text{ID} \cap \text{Isolated})}) \\ \epsilon_{\text{Bronze}} &= 1 - (\epsilon_{\text{ID}} \times \epsilon_{\text{Isolated}|\text{ID}})\end{aligned}\quad (6.4)$$

The subscript for an efficiency, e.g. $\epsilon_{\text{Prompt}|(\text{ID} \cap \text{Isolated})}$, reads as the efficiency to pass the SIP3D requirement given the lepton passes the Id and Isolation requirements. From equation ?? the Gold, Silver, and Bronze efficiencies can be read off as Gold passes all criteria, Silver fails only the SIP3D requirement, and Bronze fails either the Id or isolation and is agnostic to SIP3D. While isolation and vertexing requirements are physically uncorrelated, there is an intersection between the two, meaning a lepton can be both prompt and isolated. This intersection then demands the necessity for conditional efficiencies. The order of the conditional efficiencies is also chosen to minimize the number of measured efficiencies by reusing efficiencies across Gold, Silver, and Bronze.

Table 6.1: The criteria that define the minimum requirements for an accepted lepton. The electron and muon requirements are equivalent in terms of pseudorapidity, vertexing, and isolation but vary in p_T threshold and the MVA VLooseFO working point. The MVA VLooseFO ID also varies between years.

Criteria	Electron	Muon
p_T	$\geq 5 \text{ GeV}$	$\geq 3 \text{ GeV}$
$ \eta $	< 2.4	< 2.4
$\text{IP}_{3D}/\sigma_{\text{IP}_{3D}}$	< 8	< 8
$ d_{xy} $	$< 0.05 \text{ cm}$	$< 0.05 \text{ cm}$
$ d_z $	$< 0.1 \text{ cm}$	$< 0.1 \text{ cm}$
PFIso _{abs}	$< 20 + (300/p_T) \text{ GeV}$	$< 20 + (300/p_T) \text{ GeV}$
MVA VLooseFO ID	✓	—

The advantage of having various lepton quality categories allows for robust sensitivity to a wide range of signal processes. This strategy boosts the overall modeling statistics and

provides control regions for multiple scenarios. The populations of different truth selected objects are shown in Figure ?? and the overall efficiency for Gold, Silver, and Bronze on truth matched objects are shown in Figure ???. The gold region is mainly populated by prompt and isolated leptons that are produced within the primary vertex. This region also coincides with the signature of many targeted electroweakino models. The silver selection accommodates both leptonically decaying taus, providing an ideal region for stau's, and assists in recovering efficiency of isolated b decays in stop production. The bronze selection is rich in fake leptons and provides the best regions to extract overall fake rates for other regions as well as a surplus of events to anchor the fit.

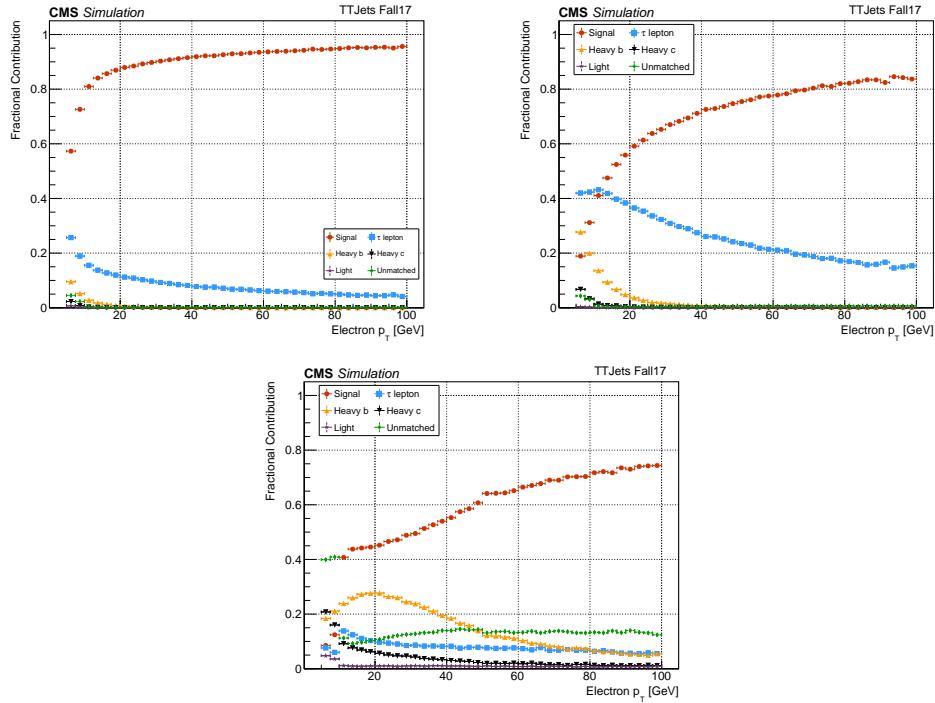


Figure 6.2: Gold (Top-Left), Silver (Top-Right) and Bronze (Bottom) MC truth matching in TTJets sample 2017. Signal is defined here as prompt electrons from a W decay.

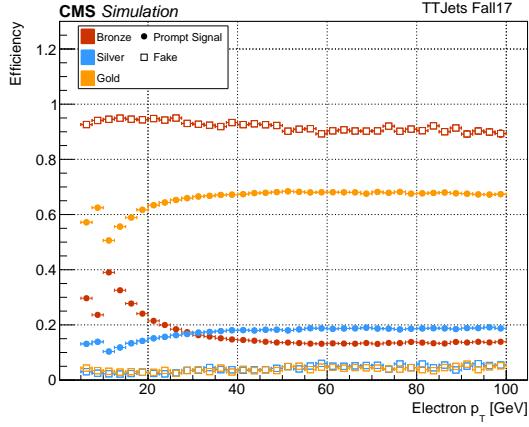


Figure 6.3: Gold, Silver, and Bronze efficiency on truth matched prompt electrons as signal and secondary electrons as Fakes.

6.3 Electron Tag-and-Probe

The electron tag and probe is done by using the Z resonance over the entire p_T range of selected electrons. The selected binnings follow the p_T and η binning conventions from the electron physics object group and are $p_T \in [5, 10, 20, 30, 40, 70, 100]$ and $|\eta| \in [0, 0.6, 1.4, 2.4]$.

The electron Tag-and-Probe tools uses a centrally curated CMSSW PhysicsTools in CMSSW_10_2_X. The software pipeline consists of two steps, an ntuplizing stage and a fitting stage. The Ntuplizing stage selects Tag-and-Probe pairs along with all potential variables of interest and loads them onto an ntuple using TnPTreeProducer [?]. The samples used in the Ntuplizing stage are listed in Table ???. In the fitting stage, a random subset of of TnP pairs are sampled with TnPTreeAnalyzer [?]. The analyzer performs all of the fitting and efficiency measurements according to the specified selection criteria.

A general selection is applied for electron TnP candidates. The selection for electrons dif-

Table 6.2: Data and MC samples for each year used for the electron Tag-and-Probe.

Type	Year	Sample Name
Data	2016	/SingleElectron/Run2016-17Jul2018_ver2-v1/MINIAOD
Data	2017	/SingleElectron/Run2017-31Mar2018-v1/MINIAOD
Data	2018	/EGamma/Run2018-17Sep2018-v2/MINIAOD
MC	2016	/DYJetsToLL_Pt-100To250_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3_ext5-v2/MINIAODSIM
MC	2017	/DYJetsToLL_Pt-100To250_TuneCP5_13TeV-amcatnloFXFX-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM
MC	2018	/DYJetsToLL_Pt-100To250_TuneCP5_13TeV-amcatnloFXFX-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM

Table 6.3: selection

Tag-and-Probe Electron Candidate Selection Criteria			
Tag	Probe	Super Cluster	Pair
$ \eta_{SC} \leq 2.1$ veto $1.4442 \leq \eta_{SC} \leq 1.566$ $p_T \geq 30.0 \text{ GeV}$ Passes Tight Id	$ \eta_{SC} \leq 2.5$ $E_{ECAL} \sin(\theta_{SC}) > 5.0 \text{ GeV}$	$ \eta < 2.5$ $E_T > 5.0 \text{ GeV}$	$50\text{GeV} < m_{ee} < 130\text{GeV}$

fers between the tag and probe, but, both depend on super cluster (SC) kinematics. The super clusters are expected to fall within the calorimeter acceptance which includes vetoing super clusters in the endcap gaps. The invariant mass of the electron of the pair also is required to fall within a specified Z-window. The selection specifics are listed in Table ???. The tag is also required to pass a trigger requirement to reflect the inherit trigger bias which is not applied in simulation by default. The triggers selected are HLT electron collections and are grouped by specific paths and filters. The electrons are matched to trigger objects in the path/filter combination and passed based on the OR of triggers in the collection. The probes are not subjected to trigger matching. The chosen trigger combinations are `HLT_Ele27_eta2p1_WPTight_Gsf_v*`, `HLT_Ele32_WPTight_Gsf_L1DoubleEG_v*`, `HLT_Ele32_WPTight_Gsf_v*` for 2016 through 2018 respectively.

The measurements of the gold silver and bronze efficiencies components, based on Equations ???, are shown in Figure ???. The relative efficiencies per component range from approximately 75% to 95% with a slight dependence on $|\eta|$ which is the strongest lower p_T . The largest combined systematic and statistical errors are $O(4\%)$ and occur in data with the lowest p_T bins. The data and MC agreement is within a few percent for both the Id and Isolation but the average data and MC agreement in SIP3D averages closer to $O(10\%)$ with the highest p_T bins discrepancies about 20% and a consistent deficit in data efficiency. The product of the efficiency components into their corresponding Gold, Silver, and Bronze category is shown in Figure ???. The efficiency for Very Loose is also included separately but is factored

into the denominator efficiencies components, so, the Gold, Silver, and Bronze efficiencies represent the overall electron efficiency for that particular lepton ranking. The range of efficiencies for each ranking are (50 – 70)% , (10 – 20)% , and (10 – 30)% for Gold, Silver, and Bronze respectively. The component combined agreement for all three ranks ranges around 10% to 20% but large discrepancies can be seen at the highest and lowest p_T bins for Silver and Bronze. Better measurements could be obtained by using a different resonance such as $J/\psi \rightarrow ee$ to measure the lower p_T ranges, however, data triggers with electrons for J/ψ are not available.

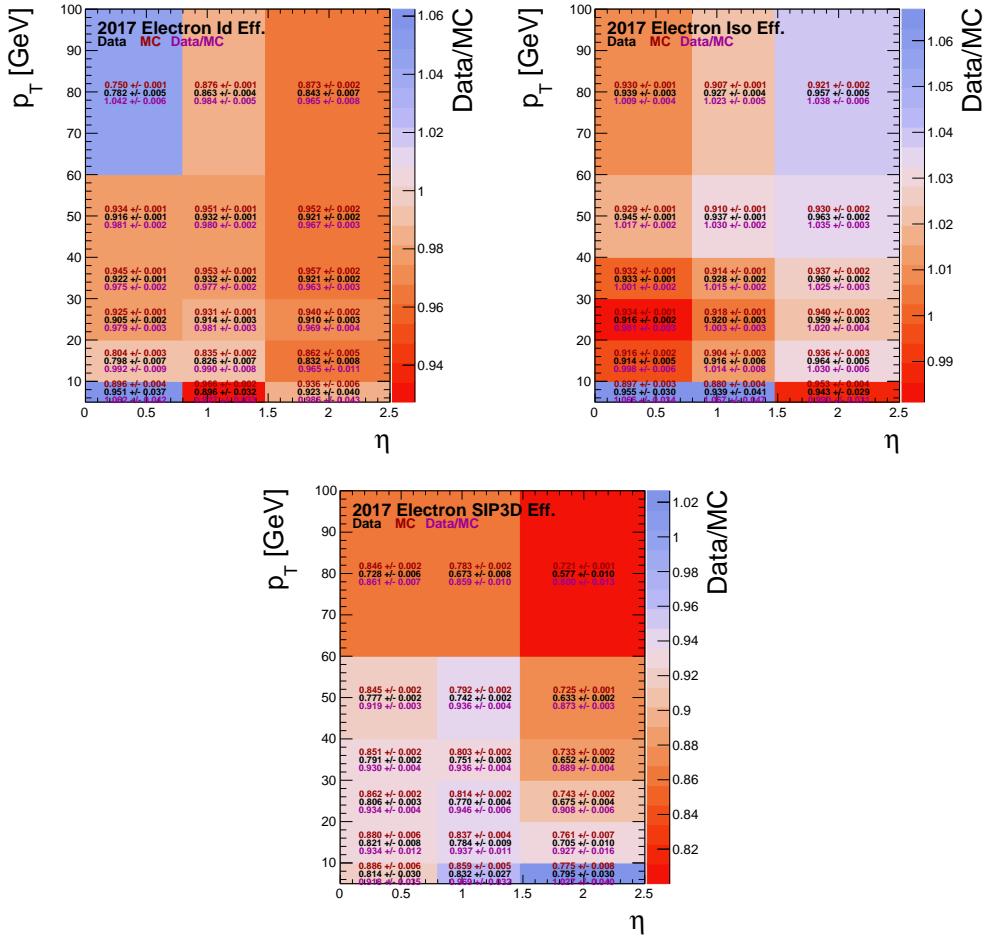


Figure 6.4: 2017 efficiencies

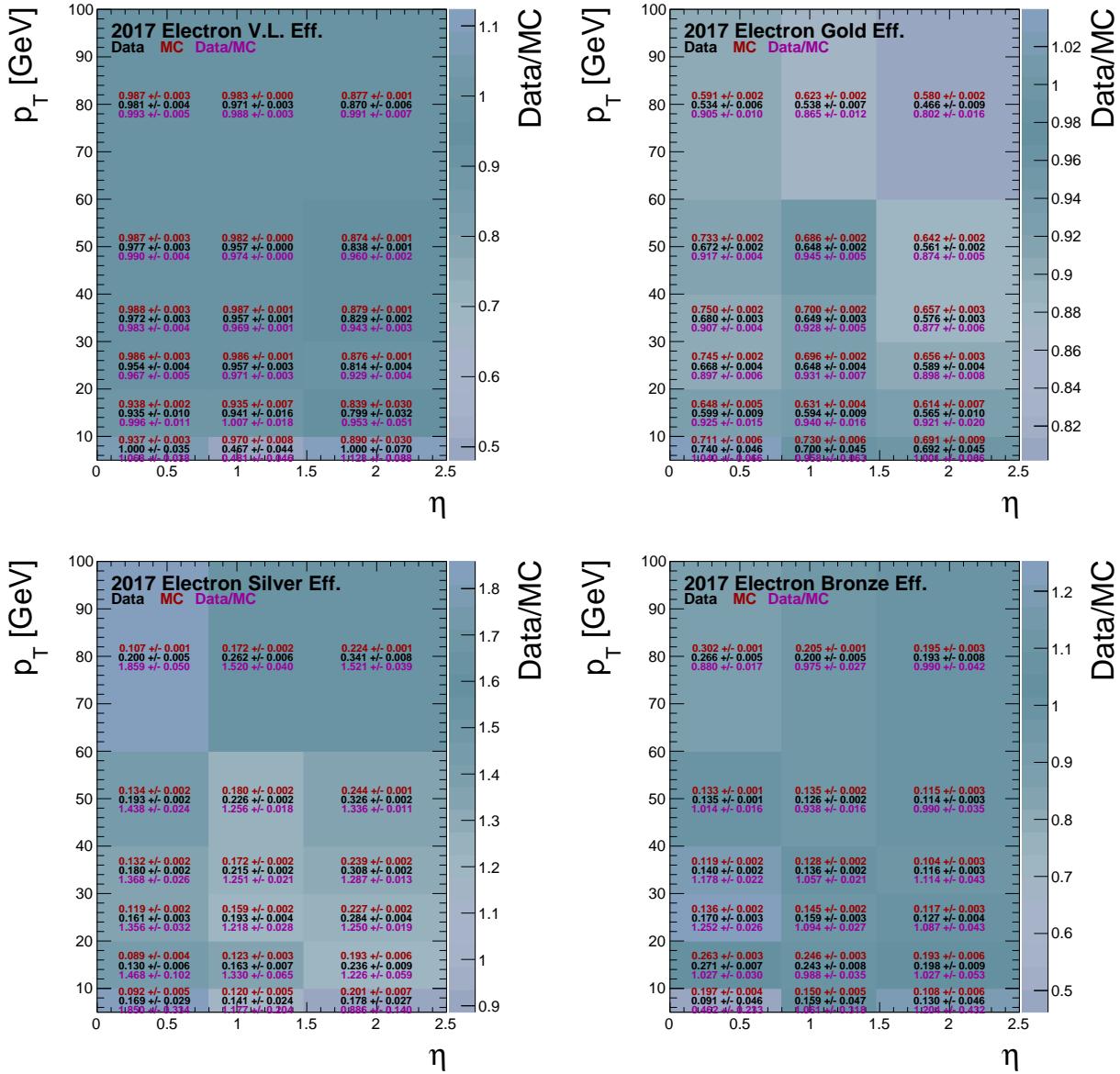


Figure 6.5: 2017 electron GSB efficiency and SF

6.4 Muon Tag-and-Probe

The muon Tag-and-Probe tools also uses a centrally curated CMSSW PhysicsTools in `CMSSW_10_6_X`. The software pipeline is identical to electrons in that it consists of an ntuplizing [?] and fitting [?] stage. The code bases for muons and electrons are separate but functionally identical. The samples chosen for Z measurements are shown in Table ???. The J/ψ ntuples are available from a central repository of standard Tag-and-Probe selection variables which use the pre-ultra legacy samples for each year [?]. The muon Tag-and-Probe efficiencies are measured above 20 GeV using the Z boson while below 20 GeV benefits from the J/ψ meson for Id measurements. The η bins are divided into a central and forward regions around the endcaps at $|\eta| = 2.1$. In total there are three sets of binnings: The low p_T J/ψ binning Z^L for muon Id below 20 GeV, the high p_T Z binning Z^H above 20 GeV, and the low p_T Z binning Z^L used to extrapolate isolation and impact parameter efficiencies down to 3 GeV. The explicit bin edges for each range are defined in Table ??.

Topological dependecies for isolation and impact parameters prevent measurement using the J/ψ . About 30% of prompt J/ψ are produced from higher mass states χ_c and $\Psi(2S)$ thus J/ψ will be produced from a cascade inside jets and likely be unisolated [?]. Similary another 10% of all J/ψ are produced within b-jets and leading to non-prompt unisolated events [?].

The exact criteria chosen for the tag and probe vary between physics processes but are identical across the two Z ranges. The selections follow the standards defined from the centrally produced muon Tag-and-Probe efficiencies.

Table 6.4

Type	Year	Sample Name
Data	2016	<code>/SingleMuon/Run2016-17Jul2018-v1/MINIAOD</code>
Data	2017	<code>/SingleMuon/Run2017-31Mar2018-v1/MINIAOD</code>
Data	2018	<code>/SingleMuon/Run2018-17Sep2018-v2/MINIAOD</code>
MC	2016	<code>/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3_ext2-v2/MINIAODSIM</code>
MC	2017	<code>/DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIIFall17MiniAODv2-PU2017RECOStep_12Apr2018_94X_mc2017_realistic_v14_ext1-v1/MINIAODSIM</code>
MC	2018	<code>/DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM</code>

Table 6.5: add ref to this table later, premade jpsi tnp trees for id

Type	Year	Sample Name
Data	2016	TnPTreeJPsi_LegacyRereco07Aug17_Charmonium_Run2016Bver2_GoldenJSON.root
Data	2017	TnPTreeJPsi_17Nov2017_Charmonium_Run2017Cv1_Full_GoldenJSON.root
Data	2018	TnPTreeJPsi_Charmonium_Run2018Dv2_GoldenJSON.root
MC	2016	TnPTreeJPsi_80X_JpsiToMuMu_JpsiPt8_Pythia8.root
MC	2017	TnPTreeJPsi_94X_JpsiToMuMu_Pythia8.root
MC	2018	TnPTreeJPsi_102XAutumn18_JpsiToMuMu_JpsiPt8_Pythia8.root

Table 6.6: muon binning

Muon Binning		
Range	p_T GeV	$ \eta $
J/ψ^L	[3.0, 4.0, 5.0, 6.0, 7.0, 9.0, 14.0, 20.0]	[0, 1.2, 2.4]
Z^H	[10, 20, 30, 40, 60, 100]	[0, 1.2, 2.4]
Z^L	[6,8,10,14,18,22,28,32,38,44,50]	[0, 1.2, 2.4]

The muon data will also have an implicit selection due to triggering. To reflect this selection in MC, the tag is required to pass a chosen trigger in the efficiency denominator in addition to HLT object matching. The triggers available vary from year to year for Z using `IsoTkMu22` in 2016 and `isoMu24eta2p1` in 2017 and 2018. A single J/ψ triggers is available for all years which is `Mu7p5Tk2`.

The Gold, Silver, and Bronze efficiency definitions are split based on p_T and reflect the high and low binning separations shown in Table ???. The low p_T muons include the Id measured by J/ψ as well as the extrapolated efficiencies from SIP3D and isolation fits in Z_L . The high p_T muons are composed of all the factors directly measured in Z_H .

$$p_T \in [3, 20)$$

Tag-and-Probe Muon Candidate Selection Criteria		
Tag	Probe	Pair
Z		
passes tightID $\sum p_T^{ch}/p_T < 0.2$ $p_T > 15$ GeV	No requirement	$m_{\mu\mu} > 60$ GeV $ z_{\mu 1} - z_{\mu 2} < 4$ cm

$$\begin{aligned}
\epsilon_{\text{Gold}} &= \epsilon_{\text{ID}}^{J/\psi} \times \epsilon_{\text{Isolated|ID}}^{Z_L} \times \epsilon_{\text{Prompt|(ID}\cap\text{Isolated)}}^{Z_L} \\
\epsilon_{\text{Silver}} &= \epsilon_{\text{ID}}^{J/\psi} \times \epsilon_{\text{Isolated|ID}}^{Z_L} \times (1 - \epsilon_{\text{Prompt|(ID}\cap\text{Isolated)}}^{Z_L}) \\
\epsilon_{\text{Bronze}} &= 1 - (\epsilon_{\text{ID}}^{J/\psi} \times \epsilon_{\text{Isolated|ID}}^{Z_L})
\end{aligned} \tag{6.5}$$

$$p_T \in [20, 100]$$

$$\begin{aligned}
\epsilon_{\text{Gold}} &= \epsilon_{\text{ID}}^{Z_H} \times \epsilon_{\text{Isolated|ID}}^{Z_H} \times \epsilon_{\text{Prompt|(ID}\cap\text{Isolated)}}^{Z_H} \\
\epsilon_{\text{Silver}} &= \epsilon_{\text{ID}}^{Z_H} \times \epsilon_{\text{Isolated|ID}}^{Z_H} \times (1 - \epsilon_{\text{Prompt|(ID}\cap\text{Isolated)}}^{Z_H}) \\
\epsilon_{\text{Bronze}} &= 1 - (\epsilon_{\text{ID}}^{Z_H} \times \epsilon_{\text{Isolated|ID}}^{Z_H})
\end{aligned} \tag{6.6}$$

The 2017 Id efficiency with statistical errors for both data and MC are shown in Figure ???. The other efficiencies for each year for all p_T ranges are included in the appendix. The overlapping bins between J/ψ and Z do not all match within statistical uncertainties. However, the average deviation of the efficiency central values are 0.02% for MC and 1% for data. The relative efficiencies per component range from approximately 88% to 98% and are fairly uniform between the central tracker and endcaps. The efficiencies for the isolation ranges from (90 – 95)% where the encaps generally are about 5% more efficient. As for SIP3D, the efficiency ranges from about (80 – 93)% with another 5% $|\eta|$ based efficiency gap, however, in the SIP3D case, the central tracks are more efficient as opposed to isolation. The extrapolation of the vertexing and isolation efficiencies below 20 GeV is done by fitting a quadratic polynomial to the efficiencies on the Z_L interval. Both data and MC are shown in Figure ???. The errors for each bin are the combined statistical and systematic errors from Table ?? and are adjusted before the polynomial fit. Any efficiencies below 20 GeV are then reported from the fit model. The fit errors are the 68% confidence interval combined with the systematic errors. The worst observed right tail P-value from all fits is $\approx 2\%$, the median P-value from the Figure ?? is 84%. The fits in each year behave qualitatively the same as 2017. The product of the efficiency components into their corresponding Gold, Silver, and Bronze category is shown in Figure ???. Similar to electrons, the efficiency for Very Loose is also included separately but is factored into the denominator efficiencies components,

so, the Gold, Silver, and Bronze efficiencies represent the overall electron efficiency for that particular lepton ranking. The range of efficiencies for each ranking are (70–80)%, (5–15)%, and (4 – 20)% for Gold, Silver, and Bronze respectively. The Data and MC agreement for all three ranks is better than electrons with the largest discrepancy in Gold being 2% and the average deviation in Silver and Bronze begin approximately (5 – 10)%.

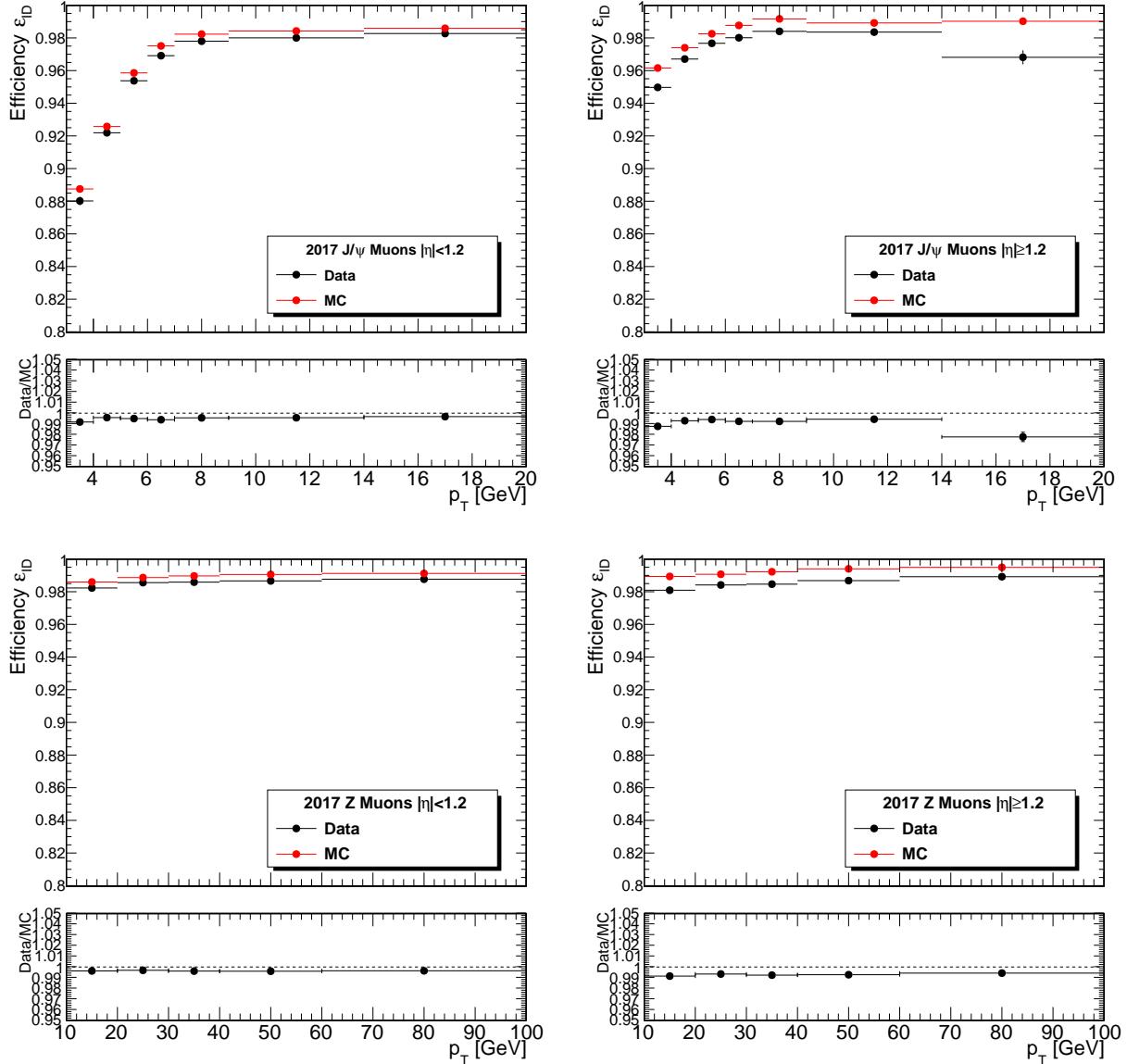


Figure 6.6: Tag-and-Probe efficiencies for the Medium Id in 2017. The left plots show the barrel while the right plots show the endcaps. The top fits use J/ψ resonance while the bottom use the Z resonance.

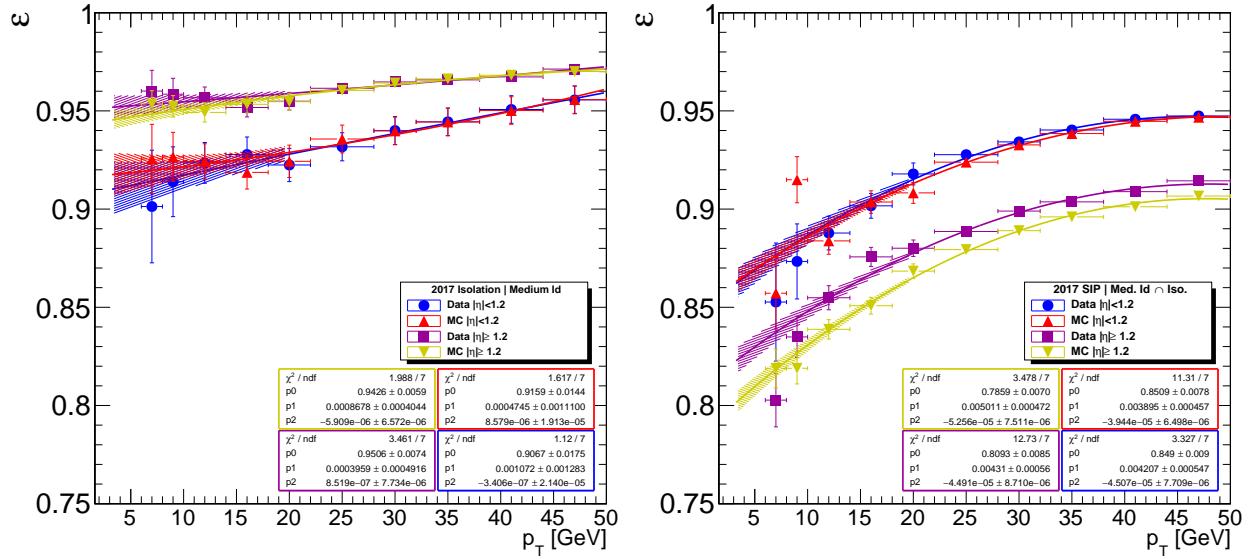


Figure 6.7: The fitted muon isolation and SIP3D efficiencies for 2017. Includes both data and MC which are separated between barrel and endcap.

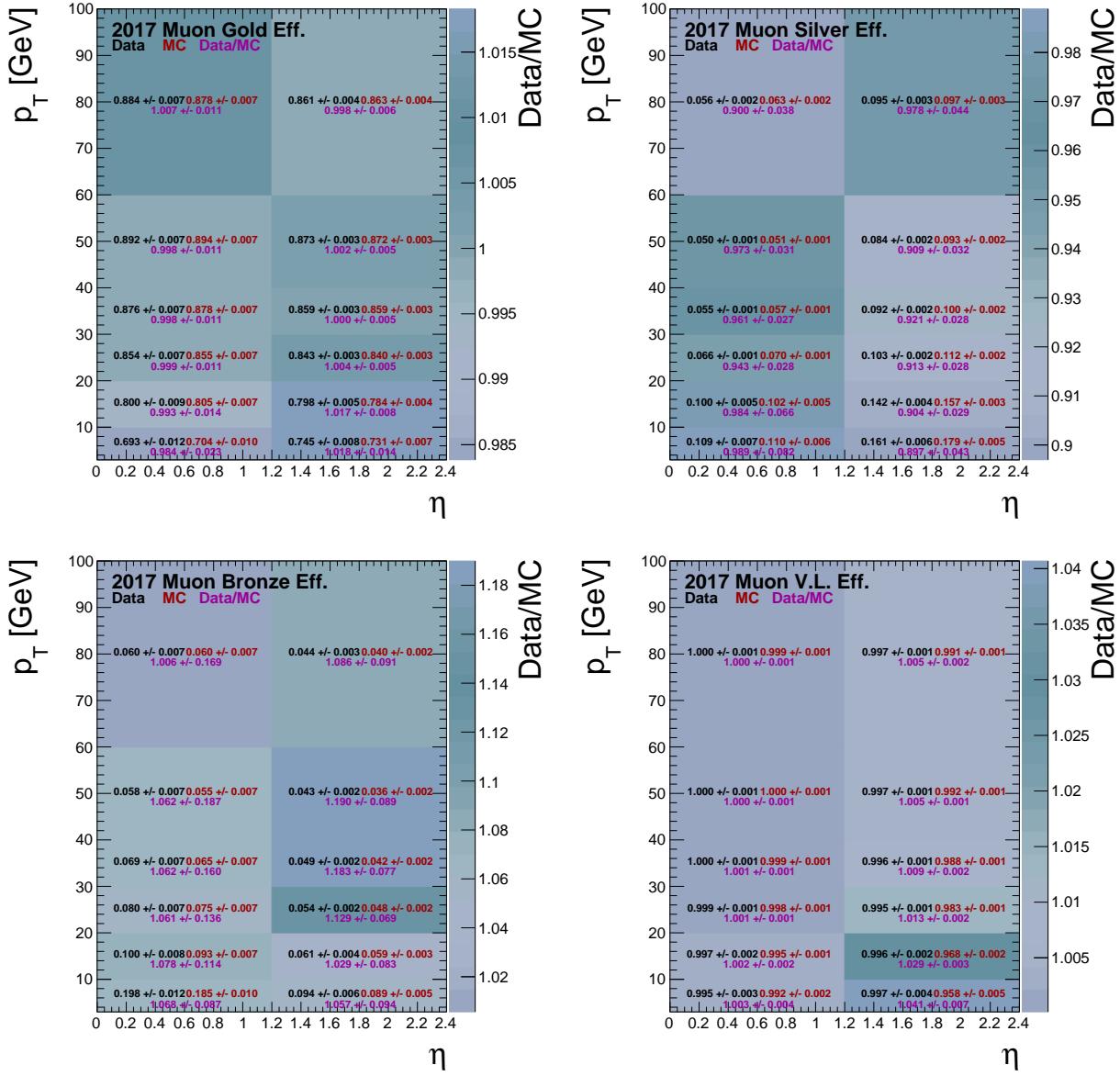


Figure 6.8: The combined efficiency components from equations ?? and ?? and Very Loose for 2017. The low- p_T region (< 20 GeV) includes the contributions from J/ψ as well as the isolation and SIP3D extrapolations. Propagated errors are treated as uncorrelated.

6.5 Lepton Systematics and Scale Factors

The systematic error for the electron and muon efficiencies are derived by varying the Tag-and-Probe signal and background models, slimming and widening the mass window , and increasing and decreasing the number of bins used in the fit. The systematic error is defined as the maximum spread in efficiencies between the modeling variations with an example spread shown in Figure ???. Rather than compute the systematic error for every bin, similarities between neighboring bins motivates using a simplified bin approach which was chosen qualitatively by the background shape. The shape of the p_T based mass distributions is illustrated in Figure ???. The same η bins are utilized according to lepton flavor, but the p_T bins are consolidated into a high and low bin pivoting on 20 GeV. A high and low systematic is derived for each selection criteria per flavor per year and is applied to the efficiencies that fall within the corresponding p_T and η range.

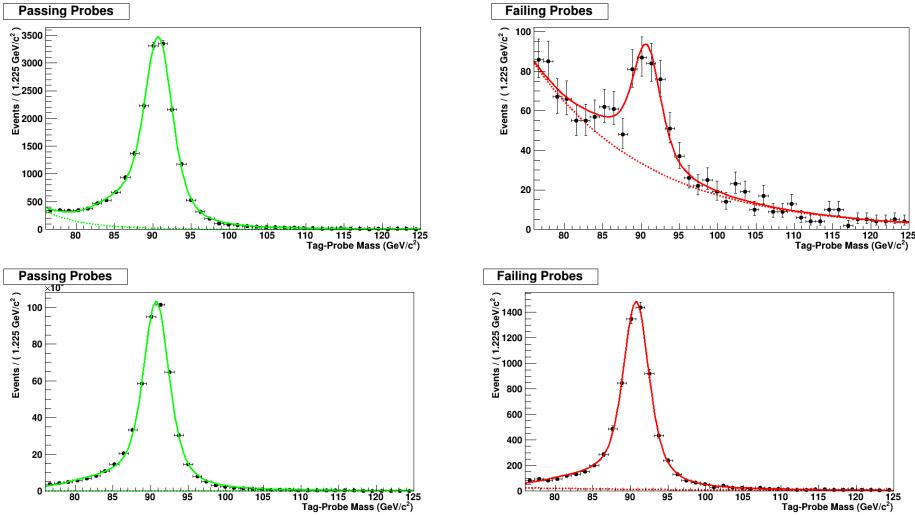


Figure 6.9: Tag-and-Probe di-muon mass distributions for both passing and failing probes. The top set of plots consist of probes below 20 GeV and the bottom set are about 20 GeV.

Scale factors are derived bin by bin for each criteria per flavor per year by finding the ratio of efficiencies in data to Monte Carlo. The scale factor variance is propagated by combining both the statistical error from the Tag-and-Probe in quadrature with the systematic error. The full 2017 set of systematics electrons and muons is shown in Table ?? and Table ??.

Additional scale factors are also needed adjusting the differences between samples which are either created with a full simulation or fast simulation. The Fast to Full factor is obtained by extracting the criteria efficiency ratio between full and fast sim ttbar samples.

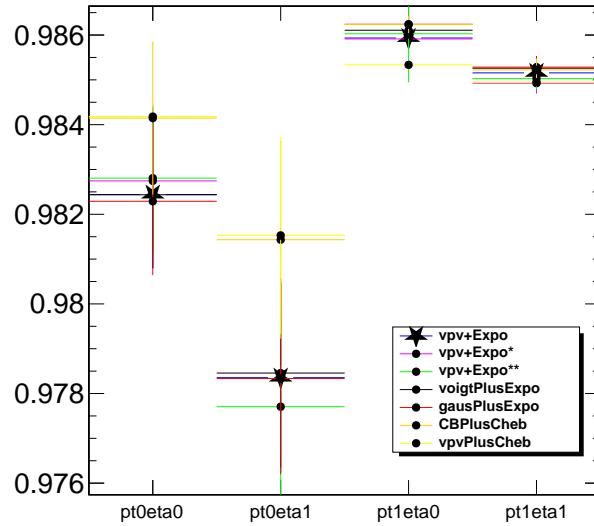


Figure 6.10: Example systematic spread from various fit models and binnings for muons. Includes the four combinations of regions either low or high pt and central and forward eta.

Table 6.7: The electron systematic error derived from the Tag-and-Probe for 2017 data and split into p_T and $|\eta|$ regions.

ID	$0 \leq \eta < 0.8$	$0.8 \leq \eta < 1.479$	$ \eta \geq 1.479$
$p_T < 20$ [GeV]	0.003	0.001	0.005
$p_T \geq 20$ [GeV]	0.001	0.001	0.002
Iso ID			
$p_T < 20$ [GeV]	0.002	0.003	0.003
$p_T \geq 20$ [GeV]	0.001	0.001	0.002
SIP Iso \cap ID			
$p_T < 20$ [GeV]	0.006	0.004	0.007
$p_T \geq 20$ [GeV]	0.002	0.002	0.0006
VeryLoose			
$p_T < 20$ [GeV]	0.002	0.007	0.03
$p_T \geq 20$ [GeV]	0.003	0.0001	0.0007

Table 6.8: The muon systematic error derived from the Tag-and-Probe data and split into p_T and $|\eta|$ regions.

ID	$ \eta < 1.2$	$ \eta \geq 1.2$
$p_T < 20$ [GeV](J)	0.001	0.001
$p_T \geq 20$ [GeV](Z)	0.001	0.0003
Iso ID		
$p_T < 20$ [GeV]	0.007	0.004
$p_T \geq 20$ [GeV]	0.007	0.002
SIP Iso \cap ID		
$p_T < 20$ [GeV]	0.005	0.003
$p_T \geq 20$ [GeV]	0.001	0.002
Very Loose		
$p_T < 20$ [GeV]	0.001	0.0003
$p_T \geq 20$ [GeV]	0.001	0.001