

A SEARCH FOR NON-RESONANT $HH \rightarrow b\bar{b}b\bar{b}$ AT $\sqrt{S} = 13$ TEV WITH THE
ATLAS DETECTOR

– OR –

$2b$, AND THEN ANOTHER $2b$... NOW THAT'S THE THESIS QUESTION.

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF PHYSICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY



Preface

There's something special about the human race that drives us not just to exist but to understand the reason for our existence. A part of the mechanistic answer to this question creates a desire to understand what are the fundamental building blocks of nature. As we moved away from the earth, fire, air, water understanding of the fundamental elements to a more modular understanding of matter's components, in 400 BC, Democritus thought that he "had it" when he coined the term "atoms" to describe what we now understand as the elements of the periodic table. Atom means "uncuttable", and is a now hilarious misnomer as the structure of the periodic table hinted at an underlying structure, experimentally verified when J.J. Thompson discovered the electrons popping off of the atom. The progression of the scientific revolution has involved us uncovering ever smaller and smaller distance scales, unfolding a fascinating and ever more intricate microcosm. In my PhD, I've been working at SLAC, one of the institutes where the charm quark was discovered, initiating the "1974 revolution" which culminated in an understanding of the parton model that helped secure the footing of the Standard Model of particle physics.

This thesis submission marks a decade since the Higgs boson discovery, and we expect to observe the SM process of two Higgs bosons with another 10 years of LHC data. Although the Higgs mechanism is now an old theory, with the massively large datasets we're collecting at the LHC, we can unlock new ways of answering these questions with the big data developments in the burgeoning field of deep learning. This thesis explores the properties of the Higgs through the lens of the Run 2 dataset in this journey of understanding the microscopic realm, This is now a quest for a whole new generation of scientists to uncover as we continue to ask ourselves the question, *What will be next?*

Summary of Contributions

The ATLAS experiment that I worked on for my PhD is a large collaboration with more than three thousand scientific authors. The physics focus of my thesis is the $HH \rightarrow 4b$ analysis, and I've also been heavily involved in optimizing the classification algorithms that identify b -jets. As these are both large group efforts, below I summarize my main contributions to each of these projects.

Flavor tagging

Recurrent Neural Network (RNN) classifier for b -tagging

- Provided the first recommendations for the RNN-based b -tagger (RNNIP) that was included in the high-level DL1r discriminant (Chapter 6.4).
 - Optimized the NN for PFlow jets extending the training to jets with p_T up to 5 TeV (Chapter 6.4.1).
 - Produced the first dedicated b -tagger trainings for variable radius track jets (Chapter 6.4.2).
 - The improvements from including the RNN classifier gave a 10% significance improvement in the $HH \rightarrow 4b$ analysis (Chapter 10.2.2).
 - This work results in both public plots [1, 2] and a paper submitted to the European Physical Journal C [3].
- Compared flipped RNN tagger definitions needed for the light-jet calibration (Chapter 6.5).
 - The recommended DL1r tagger has been now successfully calibrated for light jets with this flipped RNNIP definition [4].

R&D with Deep Sets classifier (Chapter 7)

- Applied a Deep Sets algorithm called DIPS (Deep Impact Parameter Sets) to supersede RNNIP.
- Decreased the training time by a factor of four.
- Added more low-level features and optimized the track selection to improve the light jet rejection by a factor of 2 compared to RNNIP.
- Produced ML-motivated metrics (saliency maps) to understand what the physics features the NN was learning.
- Produced both an ATLAS public note [5] and conference proceedings [6] .

$HH \rightarrow 4b$

Non-resonant Run 2 analysis

- Designed the algorithm for pairing b -jets into Higgs candidates (Chapter 10.3.2)
 - This pairing algorithm reduced the number of background events by $> 70\%$.
 - Also investigated novel attention-based transformer architectures for jet \rightarrow parton assignment.
- Proposed the categorized final discriminant for hypothesis testing (Chapter 10.5) and contributed to the statistical analysis framework.
- Proposed and organized studies for new validation regions to test the robustness of the data-driven background modeling uncertainties (Chapter 11.4).
- Internal note editor responsible for coordinating the more than 400 pages of documentation for collaboration review.
- Analysis optimization and better b -taggers improved the analysis by 30% compared to just luminosity scaling the previous analysis strategy (Chapter 12.4).
- Result public in a conference note [7], and paper submission to Phys. Rev. D [8].

Flows for $HH \rightarrow 4b$ background estimation (Chapter 13)

- Developed a novel machine learning method to interpolate into a blinded signal region.
- Implemented a neural spline flow density estimator to model smoothly varying background kinematics across the (m_{H1}, m_{H2}) massplane.
- Optimized the input variables and training procedure.
- Performed validation tests comparing to the neural reweighting method used in $HH \rightarrow 4b$.
- Paper in progress.

Acknowledgments

Every domain of human endeavor is held together by a web of relationships between people. Real people. That web is the fabric that undergirds, contains, and holds together that part of society.

– Bill Burnett & Dave Evans, *Designing Your Life*.

Can I tell you something? In the past six years, I’ve learned a lot. This thesis details what I’ve learned about HH physics and machine learning applications. But I’ve also learned how to ski, discovered that a marriage proposal from your landlord means the apartment is probably a scam, acquired a love for tequila (thanks to CV and VC), and realized that sharing drinks with strangers during a pandemic is not the best idea – and yes, it was in that order. But in the midst of all this learning, I’ve discovered something priceless: I have found a treasure in my friends, mentors, and family. I am so grateful to everyone for sharing life with me both by celebrating the successes and succoring the failures.

First and foremost, I want to thank my advisor, Michael Kagan. Michael – It’s been so exciting being your first student as you were growing your research group and program. I appreciate how you were always there for me by revising my talks, offering a fountain of ideas for things to try, and stopping everything to attend *every* big ATLAS talk I’ve given. As I got farther in my career, you didn’t need to field questions for me anymore, but I always knew you were there as my safety net in case I needed it. I appreciated how in both your actions and our interactions you taught me that academic excellence and happiness / work-life balance are not oxymorons, maybe with the best example being your establishment of weekly PGH meetings (standing for “Personal Growth and Happiness”). Thank you for taking the time to train me through reading papers together which ranged from HH to ML to b -tagging to detectors to luminosity measurements. I appreciated the interesting research directions you charted us on, and continue to strive to emulate your creative, out-of-the-box thinking. You’ve also always listened to me and been receptive to my feedback. Thank you also for all the useful feedback you’ve given me, including reading this thesis in a very short time frame.

Thank you also to my reading committee members: Su Dong, Pat Burchat, and Michael Peskin. Su Dong – you’ve always been available anytime I walked into your office to talk. One of my favorite

memories with you is the night we stayed up together doing a shift at End Station A. We chugged multiple pots of coffee together, and I felt like a “real scientist” when you had me call the control room to tell them we were ready turn on the beam. However, since then I’ve deduced from the sporadic times of your emails that the staying-up all-night part was not an uncommon event for you. Pat – thank you for being our faculty mentor in the Stanford EDGE program and giving me advice for teaching the very talented students at Stanford. My friends who have done undergrad at Stanford shared that you’ve completely transformed the student experience in the physics department. You’ve made the goals of inclusivity for students for a range of backgrounds a reality in our department, and I’m so grateful to get to be in the Stanford physics community as a result. Michael Peskin – it was fun taking QFT II with you, I was absolutely floored by how you’re a walking particle physics dictionary for any question I could think to ask (fun fact – the answers are all in [9]). I also appreciate your curiosity for understanding the details, and I learned a lot from all your questions to seminar speakers.

In terms of my physics work, Rafael’s the person that I’ve worked most closely with – and so I appreciate his guidance and mentorship. He’s a brilliant scientist, curious about a wealth of topics, and careful and methodical in each of the projects we pursued together. As a HH expert who already had produced the then world-leading κ_λ constraints on CMS in his PhD, I was delighted that he was game to stay on HH so I could also work on it in my PhD. With his wealth of expertise, we never ran out of ideas to try... just time to do them at some point! He was always available to look at any plot I was either excited about or befuddled by, and great at encouraging me to present our results, publish, and present at conferences.

The other contact for the $HH \rightarrow 4b$ analysis was the one-and-only Max Swiatlowski. He was always available and approachable to answer questions about ATLAS jargon, C++, or HH analysis design, always explaining complex concepts in an exceptionally clear way. Max has a gift for leading groups of people, and I’ve been amazed at the way he could rally our team, and appreciated all of the enthusiasm he imparted to me and the other students – all while simultaneously juggling another very time-consuming role as ATLAS jets convener! One of the highlights of my PhD was a piece of motivation he provided me for optimizing the analysis – putting a HH temporary tattoo on his face if we reached an optimization goal. Partially because of this motivation, we did reach this optimization goal, although we have yet to find a time when Max’s, the tattoo’s, and my wave functions are simultaneously overlapping.

Also, thank you to the $HH \rightarrow 4beers$ squad: Sean G, Lucas and Dale. My memories of our group meetings are infused with the excitement we had interacting with the material, as a stream of analysis-related jokes, memes, gifs, and poems would pop up in our skype chat, and then invariably infiltrate our meeting slides as well. Dale – I learned from you how to make Seinfeld references even before I had seen the show. But I must confess, I still don’t know, “*what’s the deal with JZ3???*” Thank you for making working on $4b$ an absolute delight, although sometimes I think it’s a miracle

that Max + Rafael didn't kick us out of the group.

Sean – Thank you for being my $4b$ partner in crime. You're an amazing scientist, careful researcher, prolific coder, and great at seeing the big picture, often helping me from getting caught in the details. You provided our analysis with an amazing background estimate and I appreciated that you were game to try out pairAGraph and GP interpolation studies with me. Thank you for always reading and improving my abstracts, CVs, research statements, and emails that I was nervous to send. Between doing our PhDs on the same analysis and being in the same friend group, we've ended up spending enough time together that we've ended up being able to finish each other's sentences. Thank you for giving me training labels for my jokes so I would learn how to get funnier (even though there were really only two ratings "2 / 10" and "that was actually good"). Thank you for serving as "Nicole translator" for our friends when I would start laughing so hard no one could understand my story anymore. In our friend groups, I love how you always draw others out, encouraging both jokes and warm memories, always bringing the group to a relaxed and jovial atmosphere. I'm so glad our dreams came true, and you finally became a SLACer, and it's been delightful continuing adventures in the Bay Area. In this economy, good friends are sometimes hard to come by, and I'm lucky that you're one of my best.

Lucas – you are a gem of a person, and delight to work with. I love how positive you are, careful and meticulous in your work, and how you bring up the whole environment of the group around you. Thank you for becoming the trigger expert, talking about all sorts of analysis optimizations, forming amazing collaborations, and of course, doing the illicit "cryptotuples" production with me – I was absolutely delighted how long we used those samples for.

Thank you to the other int note editors, James and Rui, for the extensive meetings to coordinate both final studies and what color coding to use as we were starting the internal review process. I think the result was the best internal note ever (or at least, the best color coordinated one). Thank you Rachel H for pushing our analysis over the finish line, as well as adventures in Croatia including both fancy food and kayaking adventures. Thank you to the other NR analyzers: Shota, for your amazing productivity from deriving the trigger SFs, to running the limits, to the Herculean effort on all of these background validation studies; Liaoshan, for your careful work and trigger expertise, Chris G for the EFT interpretations and lightning speed turn around times for code and API changes; Michael H for the ggF signal reweighting, Marta and Chris M for additional VBF expertise; Iza for likelihood scans; Todd for $3b$ category investigations, and Arely, Marco, Ebony, Marin, Thandi and Charles for validation studies. It's been so fun working with and learning from all of you this HH adventure.

Thank you also to the *amazing* FTAG community, including Chris Pollard, Francesco de Bello, Jonathan Shlomi, Manuel Guth, Bing Liu, Dan Guest, Binbin Dong, Markus Christinziani, and Geoffrey Giles. Chris, thank you for guiding me in my QT, the interesting physics discussions, and your boundless enthusiasm. Even though you had so many people who were asking for your attention

when you were FTAG convener, I loved how you always acted like you had all the time in the world whenever we talked or looked at studies together. Thank you Bing for letting me make crowns in your office any time we had a friend with a birthday, and for teaching me about FTAG calibrations. Thank you Manuel for surviving the “retraining campaign” with me – including the pain of the VR track jet collection, and becoming a wonderful friend in the process. Thank you Dan for the boundless enthusiasm and your encouragement that we needed cool names and logos for all of the taggers. While I’m sad the BUG¹ tagger never became an official ATLAS tagger, I think you were right... the name probably would have created more confusion than clarity. (“I have BUG in my derivation... do you have BUG in your derivation?” “Why would I *want* a bug in my derivation??”) Although, when so much of physics is logical and methodical, wouldn’t a tagger of chaos help keep sleep-deprived physicists on their toes? Also, thank you both Manuel and Dan for providing both technical and emotional support when I had to do basically anything C++ related. Jonathan, thank you for teaching me about both graph neural networks and Athena. Francesco, thank you for always letting me message you random questions / plots and providing detailed and thoughtful feedback and suggestions to guide me.

Through both the SLAC ATLAS group and the community of people that I met at CERN, I also met some amazing mentors who showed me what great physicists look like.

Thank you Katharine and Liza for making the ATLAS HH subgroup the most fun subgroup ever. Apparently, we have been accused of being a cult, with the community bonding traditions you developed including: doubling up our “H”s as we talked, making a special HH color template, creating holiday-themed meetings, organizing socials with virtual escape rooms, and awarding trivia winners with HH temporary tattoos and certificates of HHonour. Katharine – your physics cakes are a work of art, and I love your creativity and passion for science communication. Liza, I love your creativity with designing logos and graphic design. Both of you blended your personal enthusiasm with professional responsibilities in a way that I hope to emulate. I also appreciated how both of you were also so careful in listening to our $4b$ meetings and giving helpful advice. Thank you also to the subsequent HH conveners, Alessanda, Jason, and Valentina. Jason – thank you for editing the name of our talk to “Unbbblinded results” in the indico agenda.

Valentina – I don’t know where I shall start, as we have been all over together! From our adventures in Geneva, to HH, to your delicious and gorgeous cakes, to you helping me with job searches, to us plotting for you never to leave me in Geneva and then us plotting my path back to you, and everything in between – I can’t have imagined my time at SLAC or CERN without you. You’ve been like an older sister to me – including me getting to enjoy many of your hand-me-downs when you left Geneva. Thank you for fueling my passion to dig into the details with tracking, and spending hours answering detail-oriented questions, although some might have thought we were getting a bit dis-track-ted. Thank you for encouraging me that being a careful thinker was a positive quality when

¹The tagger BUG was named for B-jet vs Up and Gluon jet (which are light jets) discrimination.

I would get discouraged about not being a fast coder. I think one of the best compliments I ever received was when Nora said I was just like you; my heart went through the roof. PF – thank you for also teaching me about tracking and finding fun projects to work on together. I love your positivity, enthusiasm for science, and thoughtful feedback as I was deciding which next step to take in my life. I won’t know what’s behind the door I didn’t choose, but I know I’m grateful that you were there to talk through the possibilities with me. Also – thank you both for all of the support in this final writing stage.

Caterina – I am so incredibly glad that you decided to come to SLAC! Our weekly discussions about analysis and paper reading taught me how to find my own feet in the research world of particle physics. I remember the day I asked you a question, and you responded “that’s a great question, we don’t know the answer, but we can brainstorm an analysis we could do to find out.” It felt like that moment was a phase transition for me from asking questions for assimilating jargon to starting to ask research questions. Our relationship was a huge part of me feeling like I was becoming a scientist. You’ve taught me to have courage when navigating large groups, and provided the hands-on support to give me the scientific confidence to push forward. I love all the memories we’ve made through dinners and dancing, and it was also a personal relief to learn from you that very smart people can still not be the best drivers.

Ariel – I love how you are not only are a brilliant physicist brimming with creative ideas, but also a gifted communicator. I will always remember your tips for writing an engaging talk by minimizing the text, guiding my audience through the story I was telling, creating event displays whenever possible, and carefully designing plots to communicate clear and concise messages. Rafael and I enjoyed referring to your “no more than three lines per plot” maxim as “the Ariel rule.” I’ve taken this skill I’ve learned from you into every subsequent project, and for that I am truly grateful.

Charlie – thank you for guiding me into the group, and for the early direction for goals I should set for my PhD. Thank you for creating detector lectures for us as students and post-docs to make sure we were getting to see the big picture, even for the aspects that we didn’t touch on a day-to-day basis. And also, thank you for reminding me to submit the financial requests when I was nervous about asking for money. Catrin – thank you for being a great role model, always open to answering physics questions, and making physics approachable for me as a younger student both at CERN and SLAC. Eric – I love how every time you join a conversation, the whole mood lifts. Thank you for your joy and positivity.

Thank you also to the other (as in, not also on $HH \rightarrow 4b$) SLAC ATLAS post-docs.

Hannah – I’m so glad that we got to overlap at SLAC, thank you for all the time you spent with me going over application materials, and sympathizing with me over my “what should I do with my life” end-of-grad-school crisis. You always took care of us students, making us feel valued and listened to, and always provided specific help to move forward. We were so disappointed to lose you so soon – but Lund is so lucky to have such an amazing young professor. Zhi – I’m always inspired

by your passion for physics analysis, and appreciative of how you always take the time to share it with others. Thank you, Zijun, for teaching me about pixel detectors. Thank you, Brendon, for all of the positivity that you've brought to the group since joining too! Thank you, Rebecca, for leaving SLAC a better place for all of us students than when you came. I wish we had overlapped for longer. Anne of Green Gables talks about her "bosom friend", and I remember the first day we got lunch together, I had a similar feeling with a resonance between our personalities. I admire the way you work and think about problems, and appreciated your great book recommendations! Thank you also for your guidance for the next steps in life.

Furthermore, I want to thank the amazing cohort of students in the SLAC ATLAS group that I got to experience this PhD adventure with. Jannicke – although you started at Stanford a year after me, I feel like the research adventure didn't begin until you got here. Even though (I think) your favorite color is turquoise, the color that you've brought to SLAC is the warmest and brightest sunshine. I loved getting to experience all of CERN and an ATLAS PhD together, from QCD summer schools in Puerto Rico, to moving to Geneva together, enduring hikes through knee-deep snow (I don't know why), to defending our theses mere weeks apart. Also – I have yet to see someone impersonate an earthworm better than you (and yes, we did miss you so much that we did get Konstantin to try)! I love the way you're curious and creative as a scientist, and always find ways to dig in to understand the foundations of a problem. I love how you are always game to talk about anything, and there is never a silly question, just good ones! I love talking to other scientists with you because you ask the questions that I was either too shy to ask, or didn't even think to ask! Thank you for all of the encouragement in the last stage of the PhD for being brave in searching for next steps, and for always being so supportive.

Rachel S – I'm so incredibly glad you decided to come to Stanford and join the SLAC ATLAS group. A special blessing has been sharing the joys (and sorrows) of grad school through being both roommates and office mates in this final year of grad school. I continue to be amazed by your puzzle-solving skills, whether it involved playing Tetris with our porch furniture in the back of a sedan (a solution that ended up involving a piece of twine, my sweatshirt, and a hell of a lot of ingenuity), ultra-parallelizing your code racing jobs across multiple national clusters, or dumpster diving through Athena to find an exact variable definition. We've always had fun together, including making bets for how long the group meeting would go over, which resulted in us all having to put bananas in our group meeting presentations for the next year!

Murtaza – I'm so glad we got to join the SLAC group together and share memories of ML / AI classes and projects, QFT PSETs, the ins and outs of ATLAS bureaucracy, and adventures both in Geneva and Palo Alto. I remember when we met at orientation and discovered we were both rotating with ATLAS you said, "I think we'll be seeing a lot more of each other." You couldn't have been more right, and I couldn't be more grateful that you were. Eileen, thank you for all of the help and guidance with life design. Aviv, a.k.a, sensei – thank you for welcoming us younger students into

the group. Despite your vast knowledge, you never made me feel my lack of it was an impediment, and supported and taught me so much throughout my PhD. I loved our Via Ferrataing, hiking, and traveling adventures in Europe together too. Qi – thank you for mentoring me in my first rotation project and helping me prepare for opportunities beyond grad school. Dimitris – I’m so glad you decided to come to SLAC. It’s been such a joy having you in the group, thank you for helping me prepare for my defense, and all the quotes that you found for these thesis chapters. Mirella – I’m also so glad that you’re doing an ATLAS PhD and hope we can visit each other in Europe too. Sanha – thank you for all of the beautiful pictures you take at our group gatherings to save these memories for a lifetime.

I also want to thank the CS students working with Michael: Maxime and Yoann. Yoann – I’m so glad you came for your internship, and I loved both our CERN lunches and learning lindy-hop from you! Thank you, Maxime for your boundless positivity! I admire your passion and curiosity for knowledge. And hopefully we can still convince Michael to have us over for a group alumni brunch!

Thank you also to other CERN friends not previously mentioned: Danny Antrim, Sam Meehan, Kathryn Chapman, Johan, Matthew Feickert, Mason, Jennet Dickinson, Mel Buzby, Alex Armstrong, Emily Smith, Nathan Simpson, Konstantin, Aidan, and Isa. Our adventures included jumping off bridges, skinny dipping, ice skating, ariel silks classes, picnicking, hiking, sleeping in alpine huts, skiing, hostels around the continent, and during the pandemic, playing Among Us and Jackbox. In true ATLAS style, we learned to form subgroups for different social functions, so special shoutouts go to the noonch group (creatively named for the group that gets lunch at noon) and the ROCs for Jocks ML reading group.

Melissa – you are truly a diamond of a person. You’re loving, loyal, brilliant, strong, and amazingly creative. Whenever you’re hosting, it’s always a blast with whatever new game you’ve invented for us which has ranged from, Bob Ross painting parties, to gingerbread cookie portraits, to murder mysteries, and maybe most oddly with the “guess the murderer” game created from real-life 911 calls. I love how we can always just curl up together, and there’s no one I’d rather get stranded at an airport with than you. (Sure, it was a little bit embarrassing when the man next to us on the subsequent flight asked to move as we hadn’t showered in the past 48 hours. But on the plus side, it was great to have the whole row to ourselves to sprawl out in!) I loved our adventures which included butt sliding down the mountain, acro yoga, dancing and sharing the “triplet birthday” with you, Sean and me.

Thank you to the other physics students in my Stanford cohort. Kate and Clarie, I loved the memories and adventures we shared during the beginning of grad school, including surviving PSETs, learning how to ski together, and binge watching *Pride and Prejudice* (including collectively pining after Colin Firth). John – I think one of the greatest compliments I’ve ever received was when you called me a “Waltzing Machine” – and I’m so glad you decided to join Opening and become a waltzing machine yourself. To the others in my cohort, including Ronen, Connie, Tina, Abijit,

Winston, Connie, Brett, and Stan, thank you for the weekly dinners. I love our collective passion about what we do, the range of topics that we find interesting, and our attempts to morph our physics PhDs to be as ML-oriented as possible. For those in the cohort above me, Edgard thank you for all of the great lunch conversations, and fun hiking and baking adventures. Ryan – thank you for organizing the capture the flag games, and letting me beta test the video game that almost destroyed you and Edgard’s relationship. Thank you Kelly, for welcoming us all to Stanford, integrating us into the community, and leaving the whole department uplifted from you having been here. Tori, thank you for being my EDGE mentor over the years, and especially for seeing the light at the end of the tunnel as I was navigating rotations, classes, and research at the beginning of grad school. To my QFT study partners, Sean M and Conner – it was a true joy working together and it was an added bonus that our adventures in the microcosm led us all to pursue research in Europe and then meet up for some macroscopic adventures.

Thank you Christina for mentoring me in my first summer project. Although I didn’t want to keep cleaning Xenon for the rest of my PhD, I knew that I wanted to strive to become a kind and thoughtful scientist like you are during the rest of my career. Adrienne, Maria, and Nick, thank you for always keeping track of the finances and logistics that allowed us to do physics.

In the Geneva dance community, especially thank you to Lucy, Rafa, Clemance, and Malo. As COVID hit, combining zoom dance classes was such a bright spot during the wierdness of the pandemic. Thank you, Dominik for teaching me shag (don’t worry, Mom, it’s a dance style) and also the Dance Colliders group at CERN.

Also thank you to the Stanford Swingtime community including Gaurab, Jérôme, Kaya, Yao, Matias, and Micaela. Even with a love for performing and a penchant for the dangerous activities, I could not have anticipated how much I would enjoy aerals both for flipping others and for being flipped. Also, thank you for all the support over the last writing stretch. Thank you especially to Jason for being game to goof off and do aerals, still talking (let alone dancing (let alone making out)) with me after I kicked you in the face, dropped you in a trash can, and failed to spot you during your concussion. Thank you for all of the adventures, joy, and support.

Rolling back the clock a bit more, I want to thank the mentors from my undergrad at SMU who guided me in making my dream of becoming a physicist a reality. Jodi – I’m so grateful I got to be in your Modern Physics class in my first semester of school. I was so wildly homesick with my first time away from home, but you were always there to support me. When you were teaching your class, you always connected concepts from class to real-life research, and made this career path seem accessible even to undergraduates. Steve – I can’t thank you enough for welcoming me into your group, teaching me how to do research, and guiding me through my whole undergraduate career to prepare me for grad school. Words cannot thank you enough for the imprint that you’ve left on my life and career, and I’m proud to be a part of your legacy of scientists that you’ve trained. Thank you to the other incredibly supportive members of the physics department: Roberto Vega

for suggesting I try the physics major in my scholarship interview, Randy Scalise and Fred Olness for leading the SPS group, and Ryszard Stroykowski for the interesting Higgs and di-Higgs physics discussions. Thank you also to the Presidents Scholars program led by Martha Stark, Scott Norris, and Joe Kobylka. I so appreciated the breadth of a liberal arts education and how all of you helped us maximize the wealth of opportunities that SMU had to offer.

Richard Scaletter – I learned so much from you in my Summer REU, thank you for teaching and training me. You taught me so much about physics, but I feel like the biggest takeaway that I brought forward in my career was your methodical research style. The way you made lists of our open items, let us check them off together and then continually evaluate our status was really motivating for me for understanding and making progress our project. I’ve tried to emulate this approach from you in every research project I’ve done since.

Mayisha – thank you for being my “Sister in Science” and sharing both the joys of the physics degree SMU and life in grad school and beyond together.

And finally, I want to thank my family: Mom and Dad, and best siblings Brendan, Shawna, and Drew. Thank you, Brendan, for always being available for heart-to-heart phone calls, and lending me your artistic skills for making physics logos. Thank you, Shawna, for always helping me dress for success with the many gifts you’ve given to me knowingly and unknowingly with the items I “adopted” while you weren’t watching your closest. Thank you, Drew, for teaching me how to play volleyball, and encouraging me because you knew I would perform better with positive affirmation. Thank you to my cousins, Jessica and Payam, for flying all the way from the east coast for my defense, and Payam for all the deep conversations. Dad – thank you for sharing and encouraging my STEM interests, and reminding me of your mother, Ginny Hartman, who pursued math and physics degrees in the 1950s, and would have loved to pursue her own career further if it had been socially acceptable.

And of course, Mom, I love your enthusiasm to learn particle physics so you could understand the details of my work-life when I wanted to share with you, complete with extensive sets of notes you would take and then revise. I’m so impressed that you learned that “b-tagging” and “4b” were different projects, and memorized the names of all collaborators on both projects. Thank you for carrying on the legacy that your parents left us, Grandpa as a biology professor with an insatiable love of learning, and Granny Hope who spent so much time with us as kids and was such a gifted teacher. You’re not only my mom, but also my best friend. I love your joy and zest for life, and how you always can keep yourself entertained. We’ll always be a package (we belong together), but maybe since this is a physics thesis I should say we’re like the quarks in a meson, I feel like neither of us could exist alone and there’s a strong force holding us together.

Contents

Preface	iv
Summary of Contributions	v
Acknowledgments	vii
1 Introduction	1
I Foundations	4
2 The Standard Model	5
2.1 Quantum Electro Dynamics	6
2.2 The Higgs mechanism	7
2.3 Electroweak interactions with Fermions	12
2.3.1 Leptons	12
2.3.2 Quarks	13
2.4 Quantum Chromodynamics	15
2.4.1 Jets	17
2.4.2 Parton distribution functions	18
2.5 What we’ve learned about the Higgs	20
3 The Large Hadron Collider	24
3.1 The accelerator complex	25
3.2 Luminosity	25
3.3 Pileup	28
4 The ATLAS detector	29
4.1 Inner Detector	32
4.1.1 Pixel Detector	33

4.1.2	Silicon Microstrip Trackers	35
4.1.3	Transition Radiation Tracker	35
4.1.4	Tracker Resolution	36
4.2	Calorimeter	37
4.2.1	ECAL	38
4.2.2	HCAL	41
4.3	Muon Spectrometer	42
4.4	Trigger system	44
5	Event Reconstruction	45
5.1	Tracks	46
5.1.1	Track reconstruction	47
5.1.2	Challenges in Dense Environments	49
5.2	Vertexing	50
5.2.1	General problem formulation	50
5.2.2	Primary vertex reconstruction	53
5.3	Jets	54
5.3.1	Anti- k_t jet clustering algorithm	54
5.3.2	Calorimeter Jets	55
5.3.3	PFlow Jets	55
5.4	Pileup jet rejection	57
5.4.1	Boosted jets	58
5.4.2	Variable Radius track jets	59
5.5	Leptons and Photons	61
5.5.1	Electrons and photons	61
5.5.2	Muons	61
5.5.3	Taus	62
5.6	E_T^{miss}	63
II	Flavor tagging	64
6	ATLAS Run 2 b-taggers	65
6.1	Datasets	67
6.1.1	Hybrid sample creation	67
6.1.2	Sample details	68
6.1.3	Jet selection	69
6.2	Evaluating tagger performance	69
6.3	Low level taggers	71

6.3.1	IP2D and IP3D	71
6.3.2	RNNIP	75
6.3.3	SV1	79
6.3.4	JetFitter	83
6.3.5	Low-level tagger comparison	85
6.4	DL1r High Level tagger	87
6.4.1	PFlow optimization	88
6.4.2	VR track jets optimization	90
6.5	FTAG Calibrations	93
7	b-tagger R&D: DIPS	97
7.1	DIPS Algorithm	97
7.2	Baseline performance	100
7.2.1	Interpretability	101
7.2.2	Calibratability	102
7.3	Track Selection Optimization	104
7.4	A look to the future	106
III	$HH \rightarrow 4b$ analysis	109
8	Statistical techniques	110
8.1	The likelihood	111
8.2	Test statistic	111
8.2.1	Asymptotic approximation	112
8.3	Setting limits: CL_s	113
8.4	Model implementation for analyses	115
9	HH Physics overview	116
9.1	κ framework	118
9.2	Signal parametrization	121
9.2.1	ggF: histogram based reweighting	121
9.2.2	VBF: event level reweighting	124
9.3	EFT interpretations	125
9.3.1	SMEFT	125
9.3.2	HEFT	126

10 Analysis selection	129
10.1 Triggers	130
10.1.1 Trigger buckets	131
10.1.2 b -jet SF	132
10.1.3 Kinematic SF	133
10.2 Analysis jets	134
10.2.1 Jets	134
10.2.2 b -jets	135
10.2.3 b -jet corrections	135
10.3 Event reconstruction	138
10.3.1 ggF and VBF Channel orthogonalization	138
10.3.2 Higgs candidate reconstruction	138
10.3.3 Background Reduction	142
10.4 Signal Cutoffs	146
10.5 Analysis Categories	150
11 Background estimation	153
11.1 Reweighting overview	154
11.2 Background Composition	159
11.2.1 $t\bar{t}$	159
11.2.2 QCD flavor composition	161
11.3 Background systematics	162
11.3.1 Deep ensembles + bootstrap error	162
11.3.2 Choice of control region	164
11.3.3 3b1f non-closure uncertainty	168
11.3.4 Summary of Background Modeling Uncertainties	168
11.4 Background validation	171
11.4.1 Reversed $ \Delta\eta_{HH} $	171
11.4.2 Shifted regions	173
11.4.3 MC validation	177
12 Results	179
12.1 Background prediction	179
12.2 Overview of signal systematic uncertainties	181
12.3 $HH \rightarrow 4b$ analysis results	184
12.4 Combination and analysis improvements	187

IV	Future developments	189
13	Generative models for data-driven background modeling	190
13.1	Gaussian Processes	191
13.1.1	Introduction	191
13.1.2	Application for the $HH \rightarrow 4b$ analysis	193
13.2	Normalizing Flows	194
13.2.1	Intro	194
13.2.2	Implementation details	197
13.3	Results	202
14	Conclusions	208
A	Additional FTAG optimizations	210
A.1	VR track jet optimizations	210
A.2	JetFitter mass constraint	213
B	Further statistics details	218
B.1	Asymptotics approximation	218
B.2	A pedagogical motivation for the CLs test statistic	219
C	Extra analysis plots	221
C.1	Event selection	221
C.2	Cutflows	223
C.3	Background QCD composition	230
C.4	Background systematics plots	230
C.5	Pre-fit plots	232
C.6	Additional limits	234
D	3b1l category	236
D.0.1	Overview	236
D.0.2	Event selection	236
D.0.3	Background estimation	237
D.0.4	Pre-fit plots and estimated impact on the limits	239
E	Loss function derivations	242
E.1	Cross-entropy loss function	242
E.2	Reweighting loss function	243

F	$t\bar{t}$ aware reweighting	245
F.1	Motivation	245
F.2	Fitting $t\bar{t}$ templates	246
F.2.1	Prescription	246
F.2.2	Fit results 4b	248
F.2.3	Fit results 3b1f	252
F.3	Pure QCD reweighting	255
F.3.1	Mathematical formulation	255
F.3.2	Experiments	256
G	Flows optimization studies	260
G.1	Real NVP	260
G.2	flow optimizations	261
G.2.1	Modelling variables	261
G.2.2	Choice of flow model	263
G.2.3	Hyperparameter optimization	263

List of Tables

4.1	List of the dimensions of the subsystems in the ID [43, 48].	32
6.1	Pseudo-continuous (PC) working points definitions. PC bin 5 is the highest b -jet purity bin, and PC bin 1 fails all of the b -tagging WPs.	70
6.2	Categories for defining the IP2D and IP3D templates [106].	74
6.3	Track features for the RNNIP algorithm.	77
6.4	Features from the SV1 reconstruction that are fed as input to the DL1r tagger. . . .	81
6.5	Features from the JF reconstruction that are fed as input to the DL1r tagger. The first block of variables quantifies the global properties of the displaced vertices and decay topology. The second set of variables just looks at the properties of the first displaced vertex which capture the differences between b -jets and c -jets.	85
7.1	Timing metrics for trainings on an Nvidia 2080 Ti GPU and evaluations on an NVIDIA Titan X GPU. The nominal value denotes the mean of five independent trainings, while the error bar is the standard deviation. These metrics are for the inference and evaluation are for the full test datasets of 3 million jets each [5].	101
7.2	The average per jet total number of tracks (n_{trk}), the number of tracks from heavy flavour decays (n_{trk}^{HF}), the number of tracks from hadronisation, excluding those from heavy flavour decays (n_{trk}^{hadr}), and the number of tracks from mismeasurement, material interactions, and pile-up (n_{trk}^{other}), are shown for the <i>nominal</i> and <i>loose</i> selections for each jet flavor [5].	105
9.1	Cross-sections for single H and H production at $\sqrt{s} = 13$ TeV. The single Higgs cross-sections are in pb while the HH cross-sections are in fb. [137, 138]	117
9.2	Coupling values defining the basis functions for the VBF signal reweighting.	124
9.3	The coefficient values corresponding to the BSM templates, as proposed by [148] . . .	128
10.1	Triggers used for non-resonant searches. For b -tagging in the trigger in Run 2, the MV2 version of the b -tagger is used. Also, an L1 $ \eta < 3.2$ cut is assumed where not specified.	130

10.2	Luminosity by year for the $HH \rightarrow 4b$ analysis.	133
10.3	Improvements from the newly retrained taggers compared to the older tagger on the EMTopo collection. Highlighted is the improvements from the optimized tagger and WP.	135
10.4	2016-18 data yields at each step in the analysis event selection for 4b events in the ggF (top) and VBF (bottom) channels, alongside the ratio of each yield to the initial yield and to the yield for the previous cut. Highlighted are the yields in the 4b SRs.	148
10.5	Yields in the 4b ggF and VBF SRs for VBF and ggF HH signals normalized to 126.1fb^{-1}	149
11.1	Set of input variables used for the 2b to 4b reweighting for the ggF and VBF channels. The variables included in the background estimate are denoted with a checkmark.	156
11.2	Percentage of the data-driven background estimate expected for $t\bar{t}$ events.	159
11.3	Magnitude of error components for the ggF and VBF channels in the Signal Region. Total statistic error is the quadrature sum of bootstrap and 2b stat errors. The errors are just showing the impact on the overall normalization without the categorization.	168
11.4	Summary of the categorization strategy and background-related nuisance parameters.	170
11.5	Center locations for the shifted SRs validation study. The right-most column indicates which quadrants form the “nominal” background estimate.	174
11.6	4b and background prediction in the signal region in the shifted regions in 2016. The error of background prediction includes the 2b poisson statistic error, the bootstrap error and the shape systematic error.	175
11.7	(Unweighted events) in CR1 for both data and the QCD samples. Indicated in the QCD sample name is the contribution to the (weighted) event yield. The highlighted rows indicate that the JZ2 MC sample has only 1/10 of the data statistics.	177
12.1	The yields in each analysis category for the data, expected background (from the background only fit), and SM ggF and VBF signals [7].	181
12.2	Note: the (*) indicates these are only included in the μ_{SM} result.	182
12.3	Impact of the various uncertainties on the $\mu_{ggF+VBF}$ limit. To test the impact of each systematic, it’s profiled and then constrained to its best fit value as the upper limit is recalculated. Only the systematics that have an impact larger than a % are shown. All of the experimental uncertainties together had a sub-% level impact on the final result.	184
12.4	In the combined channel, the observed and expected limit intervals on the coupling modifier κ_λ at the 95% CL for the ggF channel, the VBF channel and the combination of the two.	185

12.5	The extracted upper and lower limits on the SMEFT parameters to which the analysis is sensitive. For each parameter, the constraints are provided assuming the other parameters are fixed to 0.	186
13.1	The fitted length scales for 4b data.	193
13.2	Yields for the predictions after applying the X_{Wt} cut	204
13.3	The χ^2 / (number of bins - 1) for the histograms after applying the X_{Wt} cut. In the top row, “3d discriminant” refers to the m_{HH} distribution with the $\Delta\eta_{HH}$ and X_{HH} categories.	205
C.1	ggF HH MC yields at each step in the analysis event selection for 4b events in the ggF channel normalized to 126.1fb^{-1} , alongside the ratio of each yield to the initial yield and to the yield for the previous cut.	224
C.2	ggF HH MC yields at each step in the analysis event selection for 4b events in the VBF channel normalized to 126.1fb^{-1} , alongside the ratio of each yield to the initial yield and to the yield for the previous cut.	225
C.3	VBF HH MC yields at each step in the analysis event selection for 4b events in the VBF channel normalized to 126.1fb^{-1} , alongside the ratio of each yield to the initial yield and to the yield for the previous cut.	226
C.4	VBF HH MC yields at each step in the analysis event selection for 4b events in the ggF channel normalized to 126.1fb^{-1} , alongside the ratio of each yield to the initial yield and to the yield for the previous cut.	227
C.5	$t\bar{t}$ MC yields at each step in the analysis event selection for 4b events in the ggF channel normalized to 126.1fb^{-1} , alongside the ratio of each yield to the initial yield and to the yield for the previous cut.	228
C.6	$t\bar{t}$ MC yields at each step in the analysis event selection for 4b events in the VBF channel normalized to 126.1fb^{-1} , alongside the ratio of each yield to the initial yield and to the yield for the previous cut.	229
C.7	Background flavor composition of the # of b -jets in the 4b SR in pythia QCD sample.	230
C.8	The observed and expected upper limit on the SM HH production cross-section at the 95% CL. The expected value is shown with corresponding one and two standard deviation error bounds [7].	234
D.1	Percentage of the data-driven background estimate expected to be composed of $t\bar{t}$ events for the ggF 4b (left) and 3b1l categories (right).	238
D.2	Impact on the κ_λ constraints for the various ggF categorization strategies.	241
F.1	Fitted SFs for $\alpha_{t\bar{t},2b}^{sl}$	247

F.2	Fitted normalizations for the $t\bar{t}$ and QCD templates with the pure QCD reweighting. The quoted uncertainties are from the 4b X_{Wt} shape fit.	258
G.1	Hyperparameters ranges considered for the random search.	264

List of Figures

1.1	Particles in the Standard Model of particle physics [12].	2
1.2	The goal for our physics searches: to find a model that better fits the data. Artist, Brendan Hartman.	2
2.1	The Feynman diagrams corresponding to the terms in the Lagrangian [17].	6
2.2	The form of the Higgs potential for one of the components [20].	8
2.3	Collection of measurements illustrating the running of α_S [21]. The strength of the coupling decreases as the energy, Q , increases.	16
2.4	Left: Confinement principle illustration [24]. As two quarks are pulled apart, there will be a new quark / anti-quark pair to keep each object color neutral. Right: Illustration of the formation of a jet.	17
2.5	Partons in the proton, with the valence quarks in color, and the sea quarks in white [25].	18
2.6	The proton's pdfs for $Q^2 = 10 \text{ GeV}^2$ (left) and $Q^2 = 10^4 \text{ GeV}^2$ (right) [26].	18
2.7	Event from a Monte Carlo generator [27]. The protons come in from the left and right, and the top red circle shows two gluons giving a ttH event. The lower purple blob is another parton collision coming from the same protons. The quarks and gluons fragment into other quarks and gluons. The light green blobs show the circles condensing into hadrons, while the dark green circles show the hadron decays. . . .	19
2.8	Left: The Higgs branching ratio for the leading decay modes as a function of the Higgs mass [29]. The dashed line indicates the measured Higgs mass at 125 GeV. Right: Allowed mass range for the Higgs boson at the beginning of 2012 [28]. The yellow regions are excluded by LEP and the Tevatron and LHC, while the parabolas indicate the $\Delta\chi^2 = \chi^2 - \chi_{\min}^2$ from the electroweak fits.	20
2.9	The invariant mass in the two channels used for the Higgs discovery: $H \rightarrow ZZ \rightarrow 4\ell$ [30] (left) and $H \rightarrow \gamma\gamma$ [31] (right).	21
2.10	Higgs production modes at the LHC: (a) ggF, (b) VBF, (c) VH, (d) ttH, and (e) tH [32].	21

2.11	The Higgs production (left) and decay (right) modes. In the subpanels, the vertical line delineates the processes with a 5σ observation and the process that have lower significance [32].	22
2.12	Linearity test for the strength of the coupling with the masses for the sundry production and decay channels [32].	23
3.1	Illustration of the LHC [38]	24
3.2	Cross sections for some key processes of interest for the Tevatron ($p\bar{p}$) and LHC (pp). The dashed grey line indicates the Run 2 energy, while the dotted lines indicate other past or planned collider energies [40].	26
3.3	Schedule for the LHC data taking [41]. The center of mass energy is indicated by the red horizontal lines above the timeline, while the instantaneous luminosity is shown in the red graphs below.	27
3.4	Run 2 LHC delivered and ATLAS recorded luminosity [42].	27
3.5	PU density profile for Run 2 data taking [42].	28
4.1	The ATLAS detector [43]. The people depicted on the left indicate the detector scale.	29
4.2	Schematic of different detector subsystems optimized for different particles [44].	30
4.3	Relation between η and θ [45].	31
4.4	Illustration of the subsystems in the ATLAS Inner Detector with the full layout including the endcaps (left) and just the barrel at $\eta = 0.3$ (right) [43, 47]. The IBL is not pictured here as it was installed at the end of Run 1.	32
4.5	How a silicon detector tracks incident particles [49].	33
4.6	ATLAS pixel sensors [47]. Initially they have an n ⁺ -in-n configuration (left figure), but the irradiation creates a “type-inversion” and subsequently the sensors operate in a p-in-n configuration (right figure). [47]	33
4.7	Impact of the IBL on the track’s momentum resolution, as a function of track p_T . The 2012 data is from Run 1 and does not include the IBL, while the 2015 data from Run 2 shows the lower (better) resolution with the IBL included [53].	35
4.8	Principle of operation for the TRT [54]	36
4.9	Calorimeter system at ATLAS [43].	37
4.10	The development of an electromagnetic shower [55]. The dashed lines and numbers indicate the number of radiation lengths, X_0	38
4.11	Schematic illustrating how a sampling calorimeter causes an incident particle to form a shower more quickly. [58]	38
4.12	Motivation for the accordion structure [59]	39
4.13	Wedge with the accordion structure in the LAr calorimeter [60].	40
4.14	Signal in the EM barrel calorimeter [61].	40

4.15	An element of the tile calorimeter [43].	41
4.16	Monitored Drift Tubes: MDTs	42
4.17	Muon trigger system [43].	43
4.18	Illustration of the data acquisition chain – a sequence of decisions in hardware (L1) and software (HLT) which determines which events get saved [63].	44
5.1	Illustration of what a jet looks like in our detector [64].	45
5.2	Visualization of the track helix with the perigee parameters indicated [67].	46
5.3	The track in the transverse (left) and $\rho\phi$ (right) planes. The track in the perigee representation is defined at the point of closest approach, P, relative to a reference position, R [68].	47
5.4	Graphic for the ATLAS inside-out track reconstruction steps [69].	48
5.5	Illustration of the challenges for reconstructing the charged particle trajectories in the dense environments at the core of jets [71]. Left: In the core of a jet, several charged particles can be reconstructed in a single cluster. Right: The average separation (both in the transverse, x , and longitudinal, y directions) of the closest stable particles in jets as a function of jet p_T . Also shown on the figure is the pixel size for the standard pixels and the IBL pixels.	49
5.6	Left: Fermi function (Eq. 5.8) for a range of different temperatures, $T = 1, 4, 64$. Right: Corresponding track weights for the same three temperatures [73].	52
5.7	Event display from one of the first Run 2 events. In the left image shows the (r, ϕ) view, and the colored dots show the hits in the Silicon detectors (IBL, pixel and SCT). In the lower right image, the z axis is oriented horizontally, and there are set of reconstructed vertices from a single bunch crossing [76].	53
5.8	Illustration of the shapes jets created by the anti- k_t algorithm [77].	54
5.9	Illustration of the charged energy subtraction algorithm for reconstructing the PFlow objects for two hadrons π^+ and π^- with have been reconstructed in two topo-clusters (shown in the dashed blue and black lines). The pre-sampler, three ECAL layers, and first two TileCal layers are shown, and the energy of the π^+ is in red and the energy of the π^0 is in green [62].	56
5.10	Steps of the 4-momentum calibration for PFlow jets [78].	56
5.11	Improvement due to the PFlow jet reconstruction algorithm.	57
5.12	Schematic of a hard scatter jet (in blue) and the two types of pile-up jets that we try to reject (in red and orange).	57
5.13	Illustration of the large-R jet defined from the calorimeter clusters, and the track-jets which identify the two hard prongs inside the jet. The provide a jet axis for b -tagging algorithms [83].	59

5.14	Illustration of the motivation for VR track jets The left depicts that with fixed radius track jets, for high enough boosts the products become so collimated that separate small-R jets cannot be reconstructed. The right figure illustrates how the decreasing radius with increasing sub-jet momentum can allow the separate sub-jets reconstruction of the individual b -hadron decays [83].	60
5.15	Comparison of the truth labelling efficiency for the fixed $R=0.2$ track jet, and the VR radius track jets with different choices of ρ . The truth labelling efficiency is the probability that each of the b -hadrons is within $\Delta R = 0.3$ of the sub-jet axis [83]. . .	60
5.16	Examples of hadronic τ -decays [88].	62
5.17	Set up for the hadronic τ identification. An LSTM is a type of RNN architecture. . .	62
6.1	Schematic for the characteristic “long” lifetime of a b -hadron [91].	65
6.2	Illustration of what a b -decay looks like in the ATLAS detector. The cyan colored lines illustrate the tracks from the b -hadron decay, and in the inset figure shows the displacement of these tracks from the primary vertex. Only three pixel layers are shown as this is a Run 1 event, and the IBL was not yet installed [92].	66
6.3	The p_T spectrum for training the Run 2 b -tagger recommendations (RNNIP and DL1r). . .	68
6.4	Distributions of a b -tagging discriminant as defined in Eq. 6.2 with $f_c = 0.07$, for b , c , and light-jets [5].	70
6.5	Types of b -taggers used on ATLAS	71
6.6	Lifetime signage illustration, positive for displaced decays [103].	71
6.7	Lifetime signed transverse (a) and longitudinal (b) significances for b -jets, c -jets and light-flavor jets [5].	72
6.8	Different modeling assumptions for probabilistic graphical models. A Naïve Bayes model is a generative model that assumes that the tracks’ features are <i>independent</i> [104].	73
6.9	Correlation between the two leading significance tracks in b -jets (left) and light-jets (right) [93].	74
6.10	Illustration of how an RNN works. The $X_1, X_2, X_3 \dots$ are the inputs, the $h_i, i = 1, \dots T$ is the hidden state of the RNN while the f_W is the function that updates the hidden state with the new input. The W are the weights which are common across the timesteps X_i . The sequence outputs a classification label y [108].	75
6.11	RNNIP architecture (modified from [93]).	76
6.12	The evolution of heavy flavor compared to the number of fragmentation tracks that have $p_T > 1$ GeV, $ d_0 < 1$ mm, $ z_0 \sin \theta < 1.5$ mm (the IP3D and RNNIP track selection).	77
6.13	As the b -hadron becomes increasingly boosted, it will travel farther in the decay [111] . . .	77

6.14	ROC curves for for light rejection (left) and charm rejection (right) evaluated on the 5 TeV Z' sample.	78
6.15	RNNIP performance as a function of p_T with the flat 70% WP for light rejection (left) and charm rejection (right). NN tained on 5 different samples are compared. . . .	79
6.16	Visualization of the (x,y) and (z, $\rho\phi$) 2d views of the tracks for reconstructing the secondary vertex (or vertices). The arrow on the figure indicates the jet axis, and a \star shows where each weakly decaying hadrons. The solid lines are tracks from the heavy flavor decays, while the dashed lines show the other tracks associated to the jet.	80
6.17	A selection of the SV1 features that are inputs for the high-level DL1r b -tagger [3]. .	82
6.18	Visualization of the cascade topology of $B^0 \rightarrow D^0$ decay. The axes on the plot are in mm. The hadron decays are shown with grey circles. The common B^0 / D^0 flight axis is shown in the orange line, which is initialized with the jet axis, shown by the purple dashed line. [114]	83
6.19	A selection of the JF features that are inputs for the high-level DL1r b -tagger [3]. . .	86
6.20	Comparison of the performance of the low-level taggers [3].	86
6.21	The ROC curves for light and charm rejection evaluated in $t\bar{t}$ comparing the PFlow retrained high-level taggers: MV2 (BDT-based), DL1 (NN based, RNNIP not included), and DL1r (Run 2 recommendation) [115]. The stars indicate the 77% WP for DL1r and the 70% WP for MV2.	88
6.22	The p_T dependence of the b -tagging efficiency and background rejection in $t\bar{t}$ for the fixed 77% WP for the PFlow retrained high-level taggers [115].	89
6.23	Background rejection in $t\bar{t}$ for the flat 77% WP for the PFlow retrained high-level taggers [115].	89
6.24	The p_T dependence in the 5 TeV Z' sample of the b -tagging efficiency and background rejection for the fixed 77% WP for the PFlow retrained high-level taggers (blue) compared to the old DL1r tagger in orange.	90
6.25	The ROC curves light and charm rejection comparing the VR retrained high-level taggers: MV2 (BDT-based), DL1 (NN based, RNNIP not included), and DL1r (Run 2 recommendation) [1].	90
6.26	The p_T dependence of the b -tagging efficiency and background rejection for the fixed 77% WP for the VR retrained high-level taggers [1].	91
6.27	Background rejection for the flat 77% WP for the PFlow retrained high-level taggers [1].	91
6.28	Demonstration of The grey line shows the 36.1 fb^{-1} result with fixed radius $R = 0.2$ track jets from [116]. The black line shows the expected improvement of this result expected just from the larger size of the 139 fb^{-1} dataset. The blue line shows the 139 fb^{-1} result from [117] which shows gains over the expectation from the <i>dedicated b-tagger retraining</i> for VR track jets.	92

6.29	SFs for the b and b -jets for the DL1r tagger at the 77% operating point [2].	93
6.30	$t\bar{t}$ signature for the b -jet calibration. The green cones indicate the calibration b -jets. . . .	93
6.31	$t\bar{t}$ signature for the c -jet calibration. The orange cone indicates the calibration c -jet. . . .	94
6.32	Z+jets representative diagram for the light jet calibration. The blue cone indicates the calibration light jet.	94
6.33	D_b discriminant distributions for the nominal and flipped taggers. The vertical dashed lines correspond to the discriminant requirements for 85%, 77%, 70%, and 60% inclusive b -jet efficiencies, corresponding to the efficiency benchmarks used at the analysis level. [5]	95
6.34	1 - Cumulative efficiency as a function of b -tagging discriminant for RNNIP (a) and DIPS (b). In both cases, the performance remains nearly unchanged for light-flavour jets when comparing nominal and flipped taggers, while the b -jet and c -jet efficiencies drop. [5]	96
6.35	Performance of the Z +jets light jet calibration with the DL1rFlip tagger.	96
7.1	Logo for the DIPS algorithm. The couple is featured doing a dance move called a “dip” and the tan axes overlay the impact parameter definitions. Logo artist, Brendan Hartman. . . .	97
7.2	Architecture for the DIPS algorithm. The number of hidden units in the different neural network layers correspond to the final optimized architecture.	98
7.3	The track multiplicity (black) in the jet given the cut on the maximum number of tracks cut. b , c , and light jets are inclusively considered, with the nominal track selection. The track multiplicity by the track origin is also shown for heavy flavor tracks (green), fragmentation tracks (purple), tracks from material interactions (red) and tracks from other vertices (blue).	99
7.4	Light-flavour jet rejection as a function of b -jet efficiency (left) and c -jet rejection as a function b -jet efficiency (right) of the RNNIP (green) and DIPS (purple) algorithms. The central curves and error bands show the mean and standard deviation, respectively, of the rejection at each b -jet efficiency for 5 trainings. The ratios are computed with respect to the RNNIP ROC curve. [5]	100
7.5	Saliency map for b -jets with 8 tracks. The track features are shown on the y -axis, the tracks (ordered by s_{d_0}) are listed on the x -axis. The colors in each pixel represent the gradient defined in Equation 7.2 [5].	102
7.6	D_b discriminant distributions for the nominal and flipped DIPS taggers. The vertical dashed lines correspond to the discriminant requirements for 85%, 77%, 70%, and 60% inclusive b -jet efficiencies, corresponding to the efficiency benchmarks used at analysis level. [5]	103

7.7	1 - Cumulative efficiency as a function of b -tagging discriminant for DIPS. The performance remains nearly unchanged for light-flavor jets when comparing nominal and flipped taggers, while the b -jet and c -jet efficiencies drop. [5]	103
7.8	The track multiplicity (black) in the jet given the cut on the maximum number of tracks cut. b , c , and light jets are inclusively considered, with the nominal track selection. The track multiplicity by the track origin is also shown for heavy flavor tracks (green), fragmentation tracks (purple), tracks from material interactions (red) and tracks from other vertices (blue).	104
7.9	Light-flavour jet rejection as a function of b -jet efficiency (left) and c -jet rejection as a function of b -jet efficiency (right) of the nominal DIPS setup, DIPS with <i>loose</i> track selection, and Optimized DIPS with the <i>loose</i> track selection and additional IP inputs. The central curves and error bands show the mean and standard deviation, respectively, of the rejection at each b -jet efficiency for 5 trainings. The ratios are computed with respect to the DIPS ROC curve. [5]	105
7.10	Performance plots using a fixed cut with 77% b -jet efficiency. Plots (a), (b) and (c) show the b -jet efficiency as a function of jet p_T , η and average number of proton-proton collisions per bunch crossing $\langle\mu\rangle$. Plots (d), (e) and (f) show the light-jet rejection as a function of the same quantities, while plots (g), (h) and (i) show the c -jet rejection. [5]	107
7.11	Performance plots using a requirement where the b -jet efficiency is 77% in each bin. Plots (a) and (b) show the light-jet rejection as a function of jet p_T and η , while plots (c) and (d) show the c -jet rejection as a function of the same quantities. [5]	108
8.1	Hypothesis testing cases used in particle physics. Shown are histograms for a background and signal models for an invariant mass m_X . The left image shows a data distribution for observing the signal processes, while the right shows a data distribution for setting limits on the signal strength.	110
8.2	The solid and dashed lines show the pdfs for q_1 under the $\mu' = 0$ and $\mu' = 1$ hypotheses, while the dotted line shows an observed q_1 . p_{s+b} is the probability under $f(q_1 1)$ shaded to the right of the observed q_1 , while p_b is the probability under $f(q_1 0)$ to the <i>left</i> of the observed q_1 . Then $CL_s = p_{s+b}/p_b$.	113
8.3	On the left are shown the the $f(q_\mu \mu)$, $f(q_\mu 0)$ and $q_{\mu A}$ for $\mu = 0.7$ (blue), $\mu = 1$ (red), and $\mu = 1.5$ (green). The colored dots in on the right show the CL_s for each of these μ s. The expected 95% CL upper limit on μ is indicated by the “X” The turquoise (yellow) bands show the $\pm 1\sigma$ (2σ) errors, and the the black error bars indicate the 95% CL μ_{exp} limit uncertainty.	114
8.4	Gaussian constraint graphic (description in text).	115
9.1	Impact of λ on the electroweak potential.	116

9.2	HH branching ratios.	117
9.3	The leading order gluon-gluon fusion di-Higgs production Feynman diagrams.	118
9.4	Impact of the interference of the box and triangle diagrams for ggF HH production, as a function of m_{HH} [138].	119
9.5	Impact of κ_λ variations on the ggF rate (left) and m_{HH} shape (right).	119
9.6	The three tree-level vector boson fusion di-Higgs production Feynman diagrams. A convention the matrix elements names is given in the captions of the respective diagrams.	120
9.7	VBF production cross-sections versus κ_λ (top) and κ_{2V} (bottom).	120
9.8	truth m_{HH} shape for selected VBF production signals with sm (pink), the κ_λ variations ($\kappa_\lambda=2$ in yellow and $\kappa_\lambda=10$ in turquoise), and κ_{2V} variations ($\kappa_{2V}=0$ in navy and $\kappa_{2V}=1.5$ in green).	120
9.9	Impact of the signal reweighting for the $\kappa_\lambda=10$ sample for the reconstructed m_{HH} (left) and pseudo-rapidity between the reconstructed Higgs bosons (right). The SM that we reweighted from is shown in pink, while the reweighted distribution for $\kappa_\lambda=10$ is shown in turquoise and compared to true $\kappa_\lambda=10$ distribution in navy. The error bars are statistical only.	123
9.10	Illustration of the coupling modifications in the SMEFT framework [143].	126
9.11	Illustration of the coupling modifications for the BSM (dimension 6) diagrams in the HEFT framework [146].	127
9.12	The SM and HEFT benchmark signals The benchmarks correspond to the coupling values shown in Table 9.3.	128
10.1	Flow chart for the $HH \rightarrow 4b$ analysis. The first three steps need to be passed for an event to be in either analysis category. Events passing steps (4) – (6), go in the “VBF Selection” channel, while any events failing these cuts go in the “ggF Selection” channel. (7) – (9) are the final cuts that define the VBF and ggF SR.	129
10.2	Trigger efficiencies of the 2b1j, 2b2j and combined for the MC16a/d/e corresponding to years 2016-2018 for the SM ggF $\kappa_\lambda=1$ signal. Significantly lower efficiency for 2017 2b2j compared to other years is due to tighter b-tagging requirement (40% for 2017, 60% for 2016 and 2018.	130
10.3	Trigger bucket strategy for the $HH \rightarrow 4b$ non-resonant search.	131
10.4	The bucket composition of m_{HH} for the SM ggF (left) and $\kappa_{2V}=0$ VBF (right) HH MC simulation in the 4b Signal Regions. Bucket 1 corresponds to the 2b1j trigger and Bucket 2 corresponds to the 2b2j trigger.	132
10.5	Online jet kinematic scale factors of 2b1j trigger as a function of offline jet p_T in 2017. Vertical error bars include statistical uncertainties on the data, while the green bands correspond to the quadrature sum of statistical and systematic uncertainties.	134

10.6	Illustration of the standard jet correction (blue), the μ -in-jet + p_T -reco used by the HH analyses (yellow), and a NN based regression (green) [151].	136
10.7	Comparisons of m_{H1} and m_{H2} distributions before the b -jet corrections (blue) and after the b -jet corrections (red). These distributions are fitted using Bukin function, and the peak, the peak resolution, and the relative improvement are shown in the legend.	137
10.8	Illustration of the variables to implement the ggF / VBF channel orthogonalization. The purple lines and arrows indicate the three cuts for an event to pass the VBF selection.	138
10.9	The jet selection accuracy as a function of κ_λ (left) and κ_{2V} (right).	139
10.10	Truth m_{HH} distributions for correctly and incorrectly selected jets, for the ggF (left) and VBF (right) signals.	139
10.11	The three possible pairing permutations of the four HH jets into the two Higgs candidates. The opening angles between the jets in the leading Higgs Candidate are indicated, so pair number 2 is the selected pairing.	140
10.12	The pairing accuracy as a function of κ_λ and κ_{2V} , given that the correct jets have been selected.	141
10.13	Pairing accuracy as a function of truth m_{HH} , given that the correct jets have been selected. The ggF selection accuracy is derived from the ggF SM sample, and the VBF selection accuracy is derived from the VBF $\kappa_{2V}=0$ sample.	141
10.14	The $\Delta\eta_{HH}$ distribution for HH MC and blinded 4b data. The solid purple line indicates the $\Delta\eta_{HH} < 1.5$ cut applied in the ggF selection (right). This cut is not applied for the VBF channel, but denotes the categorization.	142
10.15	X_{Wt} distributions with the 2018 dataset for the ggF (left) and VBF (right) channels. The solid purple line indicates the $X_{Wt} > 1.5$ cut.	143
10.16	Selected Higgs Candidate signal mass planes.	144
10.17	Visualization of the X_{HH} distribution for correctly and incorrectly paired events with the SM ggF (left), ggF $\kappa_\lambda=10$ (center), and $\kappa_{2V}=0$ VBF (right) analysis selections. The purple line indicates the SR defining cut.	144
10.18	The Higgs Candidate massplanes with the ggF analysis selection.	145
10.19	The Higgs Candidate massplanes with the ggF analysis selection, before applying the X_{Wt} cut.	145
10.20	$HH \rightarrow 4b$ acceptance times efficiency plots.	146
10.21	Impact of the HH channels in the combination for the resonant scalar mass m_X search [153].	147
10.22	Impact of the NR analysis selection for selected ggF signals.	147

10.23	Distributions of the categorization variables for the ggF channel. To visualize the signals they are scaled by $\alpha = 300$ and 30 for the ggF SM and $\kappa_\lambda = 10$ signals, respectively.	150
10.244b	ggF background and selected signal histograms for 2016, 2017, and 2018 with the proposed binning and categorization. To visualize the signals they are scaled by $\alpha = 300$ and 30 for the SM NR and $\kappa_\lambda=10$ signals, respectively.	151
10.25	$\Delta\eta_{HH}$ categorization variable for the VBF channel. The purple line indicates the $\Delta\eta_{HH}=1.5$ category boundary. To visualize these signals they are scaled by $\alpha = 1250$ and 20, and 20 for the VBF SM, $\kappa_\lambda = 10$, and $\kappa_{2V}= 0$ signals, respectively.	152
10.264b	VBF background and selected signal histograms for 2016, 2017, and 2018 with the proposed binning and categorization. To visualize these signals they are scaled by $\alpha = 1250$ and 20, and 20 for the VBF SM, $\kappa_\lambda= 10$, and $\kappa_{2V}=0$ signals, respectively.	152
11.1	The number of events per bin as a function of di-Higgs candidate mass for the data and the Monte Carlo estimates for multi-jet and top pair-production process with 4 b -jets in Control Region 1 (defined in Chapter 10.3.3).	153
11.2	The regions in the 4b analysis.	154
11.3	Distributions of ΔR between the closest Higgs Candidate jets, ΔR between the other two (training variables) and the invariant mass of the di-Higgs system (non-training variable) before (left column) and after (right column) the CR 1 derived reweighting for the 2018 Control Region 1.	157
11.4	Distributions of the top veto variable, X_{Wt} , the energy of the leading Higgs candidate (training variables), and the invariant mass of the di-Higgs system (non-training variable) before (left column) and after (right column)CR 1 derived reweighting for the VBF Control Region 1.	158
11.5	Distributions in CR1 after the reweighting and $\Delta\eta_{HH}$ categorization for VBF. Bootstrap and Poisson errors are included.	159
11.6	Distributions in CR1 after reweighting and $\Delta\eta_{HH}$ categorization for each year of the ggF background estimate, 2016 (top), 2017 (middle), and 2018 (bottom). Bootstrap and Poisson errors are included.	160
11.7	The flavor composition from Pythia QCD of the leading four p_T b -jets used to reconstruct the HH system (“analysis jets”) in the ggF 4b SR, as a function of m_{HH} . The left plots shows the m_{HH} distribution while the right plot normalizes each m_{HH} bin to unity.	161

11.8	One dimensional non-convex, multi-modal loss function. The turquoise dots are randomly initialized weights w_i , and the black arrows are the gradients updates that minimize the loss. Each of the local minima (in the pink dots) is a different parametrized function, and the variance of these functions' predictions is the deep ensembles error. Modified from [161].	162
11.9	Illustration of the deep ensembles error, shown as a relative error with respect to the nominal estimate. Each grey line is from the m_{HH} prediction for a single NN training, and the solid pink line shows the standard deviation of the gray histograms.	163
11.10	Illustration of the derivation of the shape systematic. The top row shows the derivation of the nominal estimate from the CR1 regions, and the bottom row shows the alternative estimate derivation from the CR1 region. The ratio between these predictions is the "CR12 shape systematic". The symmetrized alternative model is shown in the pink variation histogram on the right plot. These templates are defined for each of the $\Delta\eta_{HH}$ (+ X_{HH} for ggF) categories.	164
11.11	SR quadrants chosen to derive the four background variation NPs.	165
11.12	Templates for the CR12 extrapolation systematic for the 2018 ggF discriminant in the SR NP quadrants for Q_N , Q_W , Q_S , Q_E (ordered top to bottom). The nominal histogram is shown in navy, while the alternative histograms defining the $\pm 1\sigma$ variations are shown in the pink and turquoise histograms. The ratio panel shows the ratio of these variation histograms to the nominal.	166
11.13	Impact of background shape nuisance parameter variation on m_{HH} in different kinematic categories for the ggF channel. Each column is a different year for the ggF channel templates while the rows show the SR NP quadrants.	167
11.14	Relative error contributions of the background for the ggF discriminant for the 2016 (top), 2017 (middle) and 2018 (bottom) background estimates.	169
11.15	Relative error contributions of the background for the VBF discriminant.	170
11.16	m_{HH} and $\Delta\eta_{HH}$ distributions for the rev $\Delta\eta_{HH}$ 2018 ggF data sample.	171
11.17	The ggF 4b rev $\Delta\eta_{HH}$ discriminants for the 2016 (top), 2017 (middle) and 2018 (bottom) background estimates.	172
11.18	The shifted regions for the background validation, with the pink solid curve in the center showing the standard SR. The SRs are shown in the solid lines, and the corresponding control for deriving the background estimate and CR12 extrapolation error are shown in the dashed lines.	173
11.19	m_{HH} distributions of reweighted 2b data and 4b data in the shifted SRs for the 2018 background estimates. The background error bar includes the 2b Poisson, bootstrap+deep ensembles, and the CR1 / CR2 extrapolation systematics.	174

11.20	μ_{norm} distribution in the shifted regions. 4b normalizations (black) and the gaussian fit (red) are shown.	175
11.21	Pull plots for fits with background nuisance parameters floating for the five shifted regions. The signal strength is fixed to $\mu = 0$	176
11.22	m_{HH} data reweighted events evaluated on the QCD and $t\bar{t}$ MC samples. The background error includes the deep ensembles, 2b stat, and CR1/CR2 systematic errors.	177
11.23	m_{HH} MC reweighted 2b events and 4b events for the QCD and $t\bar{t}$ MC samples. The background error includes the deep ensembles, 2b stat, and CR1/CR2 systematic errors.	178
12.1	The post-fit background model for the ggF selection, with the predictions from all years combined. The post-fit error is shown in the dashed gray lines. The 4b data is shown in black, and the sub-panels show good agreement with the post-fit background prediction. Also shown are the ggF signals for the SM (scaled by a factor of 200) and $\kappa_\lambda=6$ (scaled by a factor of 100). Each plot is a separate kinematic category with the top row showing $X_{HH} < 0.95$, the bottom row showing $X_{HH} > 0.95$, and the columns showing the $\Delta\eta_{HH}$ categories.	180
12.2	The post-fit background model for the VBF selection, with the post-fit error shown in the dashed gray lines. The 4b data is shown in black, and the sub-panels show good agreement with the post-fit background prediction. Also shown are the VBF signals for the SM (scaled by a factor of 1000), $\kappa_\lambda=6$ (scaled by a factor of 50), and $\kappa_{2V}=0$ (scaled by a factor of 10). Each plot is a separate kinematic category with $\Delta\eta_{HH} < 1.5$ on the left and $\Delta\eta_{HH} > 1.5$ on the right.	180
12.3	Pulls for the fit to the background template.	183
12.4	The 95% Confidence Level upper limits on the HH production cross-sections with respect to the κ_λ variations (left) and the κ_{2V} variations (right).	185
12.5	The observed (solid) and expected (dashed) limit intervals on the coupling modifiers κ_λ vs κ_{2V} (left) and κ_V vs κ_{2V} (right) at the 95% CL for the combination of the VBF+ggF channels. The solid lines show the couplings excluded by the observed 95% CL limits, while the dashed lines indicate the expected 95% CL limits. The blue (yellow) bands give the errors on the expected limits. The shaded grey regions are excluded. The stars indicate the SM couplings ($\kappa = 1$).	186
12.6	The extracted upper limits for the SM cross section, and the seven HEFT benchmarks.	187
12.7	The SM limit combining the ATLAS channels [153] [163]	187
12.8	Run 2 publications of the expected 4b non-resonant limit, in units of the $HH \rightarrow 4b$ cross section [164], [165], [116], [163]. For a fair comparison across the publications, the theoretical cross-section uncertainties in [163], as these were not included in the earlier publications.	188

13.1	Illustration of the interpolation setup. The pink circle shows the nominal 4b SR, with the purple dotted line showing the bounding box used to train the interpolation. The quadrants for deriving the baseline reweighing method are shown in the navy blue dashed line crescents.	191
13.2	GP fits for the 2018 4b SR. All the massplane fits are before the X_{Wt} cut.	194
13.3	Left: non-closure of the GP fits for the 2018 4b SR region. Right: 1d histogram of the pulls. The “ \neg SR” (“not SR”) purple line shows the bins used to fit the GP, while the “SR” pink line shows the bins that were blinded in the fit. These evaluation plots are before the X_{Wt} cut.	194
13.4	Visualization of the normalizing flow paradigm. The forward mode ($f = f_1 \circ \dots \circ f_L$) generates samples. The reverse mode ($f^{-1} = f_L^{-1} \circ \dots \circ f_1^{-1}$) evaluates the density of a data point.	195
13.5	RQ-NSF parametrization	196
13.6	Motivation for Higgs Candidate variables modeled by the flow. The flow models the Higgs Candidates’ p_{Ts} (a) and ηs (b), the $\log(\pi - \Delta\phi_{HH})$ (e), and X_{Wt} (not shown here).	197
13.7	Demonstration of how the flow transforms unstructured noise into a structured prediction in the 4b SR. These samples are conditioned on the SR (m_{H1}, m_{H2}, yr) data from the 2016, 2017 and 2018 datasets. The grey histograms in the left column are samples from a 6d Gaussian. Each column to the right shows the transformation from one layer of the flow (i.e, a single invertible transformation), with the right-most column showing the prediction of the final flow. This is a single flow training, with the distributions are shown before applying the X_{Wt} . Variables have been scaled to zero mean and unit variance.	199
13.8	The correlation between the modeling variables for the SR data (grey) and the flow prediction (pink) in the 4b SR. The blue box emphasizes the η_{H1}, η_{H2} variables to show how the flow has learnt the $\Delta\eta_{HH} < 1.5$ cut. The marginal plots along the diagonal are normalized to unity. These samples are conditioned on the SR (m_{H1}, m_{H2}, yr) data from the 2016, 2017 and 2018 datasets. This is a single flow training, with the distributions shown before applying the X_{Wt} cut. Variables are scaled to zero mean and unit variance.	200
13.9	Demonstration of the background prediction algorithm for the upper right SR. . . .	201
13.10	The flow training variables in the 4b SR (after the X_{Wt} cut).	202
13.11	High level variables reconstructed from HC kinematics in the 4b SR (after the X_{Wt} cut).	203
13.12	High dimensional discriminant in the 4b SR, after the X_{Wt} cut.	203

13.13	Illustration of the interpolation regimes. The pink (solid) circle shows the nominal 4b SR, with the pink dotted line showing the interpolation bounding box. The quadrants used to define the reweighing are also shown in the pink dashed crescents. The shifted regions used as validation tests of the method are shown in the blue, orange, green, and purple overlays.	204
13.14	m_{HH} for the background validation regions, shown after the X_{Wt} cut.	206
A.1	Comparison of the PFlow and VR track jet p_T for jet reconstruction.	210
A.2	The VR track jet p_T spectrum for the hybrid sample with the 125 GeV light and c -jet p_T cut.	211
A.3	Performance for trainings on four different samples for the VR track jet collection. The evaluations are done both on $t\bar{t}$ (left) and Z' sample (right). The red arrows on the plots emphasize the improved performance when the hybrid definition jet p_T cut was changed from 250 GeV to 125 GeV. The subpanels show the ratio of the performance with respect to the $t\bar{t}$ training.	211
A.4	Improvement in the performance from the dedicated VR track jet trainings (purple). The comparisons are to two previous calo-jet trainings: the old training EMTopo jets (orange) and the improved PFlow training (pink).	212
A.5	$p_T^{frac} = p_T^{trk}/p_T^{jet}$ and the $\Delta R(trk, jet)$ in $t\bar{t}$ – the RNNIP inputs with the largest difference between the PFlow and VR track jet collections.	213
A.6	Illustration of where the improvements of the dedicated PFlow training are coming from (for the light-jet rejection). The pink line is PFlow recommendation, while the cyan line is the same NN, but modifying the p_T^{frac} inputs $p_T^{frac} \Rightarrow \frac{1}{2}p_T^{frac}$. This is compared to a VR $t\bar{t}$ track jet training (left) and Z' track jet training (right).	214
A.7	Reconstruction of the B -hadron (left) and D -hadron (right) masses from the truth charged particles.	215
A.8	Comparison of the ROC curves from the JF corrected mass whether the scalar sum for p_\perp is used (solid) or the vector sum for p_\perp is used (dashed).	216
A.9	Reconstruction of the B (left) and D (right) hadron masses from the truth charged particles. The solid lines show the masses reconstructed from the scalar sum, while the dashed lines show the reconstruction from the vector sum. The purple (blue) lines show the result for randomly dropping one (two) of the truth particle(s). The roc curve demonstrates that as we drop more tracks,	217
B.1	Comparison of the pdfs reconstructed from test statistic samples (red and blue dashed histograms) and the result from the asymptotic formula (solid black lines) [132]	219
B.2	A comparison of the definitions for CL_{s+b} , CL_b , and $CL_s = CL_{s+b}/CL_b$ associated to a model $n = s + b$ [181].	220

C.1	The pairing accuracy as a function of κ_λ and κ_{2V}	221
C.2	The pairing accuracy as a function of κ_λ and κ_{2V}	222
C.3	Impact of the X_{wt} cut for the VBF HC massplanes. The top row shows the $4b$ distributions before (left) and after (right) the X_{wt} cut, and the bottom row shows the same for $2b$	222
C.4	Impact of background shape nuisance parameter variation on m_{HH} in different kinematic categories for the ggF channel. Each column is a different years of the ggF channel templates while the rows show each category with the SR NP quadrants overlaid.	231
C.5	$4b$ VBF background and selected signal histograms for 2016, 2017, and 2018 with the proposed binning and categorization. To visualize these signals they are scaled by $\alpha = 839$ and 13, and 12 for the VBF SM, $\kappa_\lambda = 10$, and $\kappa_{2V}=0$ signals, respectively.	232
C.6	$4b$ ggF background and selected signal histograms for 2016 (top), 2017 (middle), and 2018 (bottom) with the categorization. To visualize the signals they are scaled by $\alpha = 300$ and 30 for the ggF SM NR and $\kappa_\lambda=10$ signals, respectively.	233
C.8	Comparison of $b\bar{b}\gamma\gamma$, $b\bar{b}\tau\tau$, and $4b$ channels	234
C.7	Upper limits for the $b\bar{b}b\bar{b}$, $b\bar{b}\gamma\gamma$, and $b\bar{b}\tau^+\tau^-$ analysis channels and their combination [163].	235
D.1	The jet selection accuracy with the 3b1l category included.	237
D.2	Visualization of the $\Delta\eta_{HH}$ and X_{Wt} cuts for the 3b1l channel with the ggF selection.	237
D.3	3b1l Signal Region definition showing the SM signal and (blinded) background distribution.	238
D.4	CR1 after reweighting for the ggF 3b1l.	238
D.5	Distributions of the variables used for categorization in $4b$ and $3b1l$ ggF. Years are merged. To visualize the signals they are scaled by $\alpha = 300$ and 30 for the SM NR and $\kappa_\lambda = 10$ signals, respectively.	239
D.6	$3b + 1$ loose ggF background and selected signal histograms for 2016, 2017, and 2018 with the proposed binning and categorization. To visualize the signals they are scaled by $\alpha = 300$ and 30 for the SM NR and $\kappa_\lambda = 10$ signals, respectively.	240
D.7	The κ_λ scan with the ggF selection including the $4b$ and $3b + 1$ loose selections and the $\Delta\eta_{HH}$ and X_{HH} categorization (left). Impact of the extra 3b1l category compared to just the ggF categorization (right). The background shape NPs are treated as uncorrelated across the X_{HH} and $\Delta\eta_{HH}$ categories.	241

F.1	Performance of the inclusively trained reweighting evaluated on the $t\bar{t}$ simulation. The performance of the inclusively trained reweighting is evaluated on $2b\ t\bar{t}$ simulation and compared to the $4b\ t\bar{t}$ prediction. The error bar on the background prediction shows the quadrature sum of the 2b Poisson, deep ensembles, and CR1 / CR2 shape systematic error.	245
F.2	Illustration of how the data-trained reweighting (in pink) inclusively reweights all of the background components. The only background component that we need to use a data-driven technique for is the QCD piece where we don't trust our simulation. . .	246
F.3	m_{HH} in $2b$ sample with an isolated muon for the 2016 data in CR1. The dashed red line in the subpanel shows the fitted $\alpha_{t\bar{t},2b}^{sl}$	247
F.4	The first bin shows 4b events with an isolated muon and the rest of the distribution shows the X_{Wt} histogram. Overflow events are included in the highest bin. The left distribution shows the pre-fit templates and the right distribution shows the post-fit templates with optimized normalization. The statistical errors of the fit are shown in the hatched grey lines on the right subpanel. $X_{Wt} = 1.5$ (analysis cut) is also shown in the figures.	248
F.5	The correlation between the fitted template normalizations from the combined 4b fit in CR1. The columns correspond to the different years of data taking.	249
F.6	Evaluating an uncertainty based on the 4b normalization fits in CR1 and CR2. . . .	249
F.7	CR1 fit in CR1. The top rows show the pre-fit templates and the bottom rows show the post-fit norms for the templates. The statistical errors of the fit are shown in the hatched grey lines on the right sub-panel. The pink line shows the inclusively trained and evaluated reweighting (the baseline analysis strategy).	250
F.8	CR1 evaluation using the template normalizations from the CR1 fits with the $X_{Wt} > 1.5$ cut. The first subpanel shows the ratio of the 4b to the background prediction, with the black dots showing the ratio to the template fits, and the pink curve showing the ratio to the inclusive reweighting prediction. The bottom panel compares the ratio of the inclusive reweighting to the template fit prediction.	251
F.9	CR2 evaluation using the template normalizations from the CR1 fits. The first subpanel shows the ratio of the 4b to the background prediction, with the black dots showing the ratio to the template fits, and the pink curve showing the ratio to the inclusive reweighting prediction. The bottom panel compares the ratio of the inclusive reweighting to the template fit prediction.	251

F.10	CR2 evaluation using the template normalizations from the CR1 fits with the $X_{Wt} > 1.5$ cut. The first subpanel shows the ratio of the 4b to the background prediction, with the black dots showing the ratio to the template fits, and the pink curve showing the ratio to the inclusive reweighting prediction. The bottom panel compares the ratio of the inclusive reweighting to the template fit prediction.	252
F.11	Evaluating an uncertainty based on the 3b1f normalization fits in CR1 and CR2. . .	253
F.12	CR1 fit in CR1	253
F.13	3b1f SR evaluation using the CR1 fits	254
F.14	3b1f SR evaluation using the CR1 fits	254
F.15	$R_{2b^{t\bar{t}} \rightarrow 4b^{t\bar{t}}}$: The MC based $2b \rightarrow 4b \ t\bar{t}$ reweighting.	257
F.16	$R_{2b^{t\bar{t}} \rightarrow 4b^{t\bar{t}}}$: Reweighting $2b$ data $\rightarrow 2b \ t\bar{t}$	257
F.17	X_{Wt} distributions in the 4b CR1 with the pre-fit (left) and post-fit (right) plots, compared to the 4b CR1 data. The uncertainties in the subpanels of the right column are from the 4b X_{Wt} shape fit.	259
G.1	Graphic for explaining how a Real-NVP works for a 2 variable flow.	260
G.2	The extension of the Real-NVP for a conditional 2 variable flow.	261
G.3	Comparison of modeling HH kinematics versus the HC 3-vectors.	262
G.4	263
G.5	To choose the optimal model we found that minimizing the validation loss was a better metric than minimizing the SR loss.	264
G.6	Illustration of a subset of the models considered in the hyperparameter scan. The orange and blue show the training and validation loss in the interpolation bounding box (with the SR), and the pink shows the negative log likelihood of the SR data. Each model configuration has been trained 25 times, and the mean and standard deviation are plotted. The star (and yellow circle) shows the model configuration that we used for the results in Chapter A.1.	265

1

Introduction

Until taught by prolonged exposure that the universe contained anomalous cards, they saw only the types of cards for which previous experience had equipped them. Yet once experience had provided the requisite additional categories, they were able to see all anomalous cards on the first inspection long enough to permit any identification at all . . .

– Thomas Kuhn, *The Structure of Scientific Revolutions*

Particle physics is the study of the fundamental particles of nature and their interactions. The Standard Model (SM) encapsulates our current understanding of these fundamental particles, shown in Figure 1.1. Much like the periodic table organizes the properties of the sundry atoms given by the electron configurations, Figure 1.1 similarly organizes the particles with similar properties together. There are two types of particles: the quarks and leptons are *fermions* (or spin $\frac{1}{2}$ particles) that make up the matter in the universe. The gauge bosons are the spin 1 particles that mediate these forces: the familiar photon mediates electromagnetism, the gluon mediates the strong force which holds both nucleons and the nucleus together, and the W and Z bosons mediate the weak force responsible for radioactive decay. We now understand the electromagnetic and weak forces as coming from an underlying *electroweak* theory, and the process of electroweak symmetry breaking gives mass to the W and Z bosons. A prediction of this process is a scalar Higgs boson, and its discovery in 2012 was the SM particle to be experimentally discovered [10, 11],

The SM has been developed and tested through a suite of experiments in the past half a century, and thus far it has been an enormously successful theory with precise tests for a range of different processes, including an impressive suite of analyses from the last decade probing the Higgs sector.

Despite being a successfully predictive theory, there are some aspects (which Thomas Kuhn would call “anomalous cards”) that the SM can currently not explain, such as: what the nature of dark matter is, why the SM fermions are organized into three generations, whether there is any underlying structure to the masses and couplings of the SM, and whether a consistent model can also incorporate

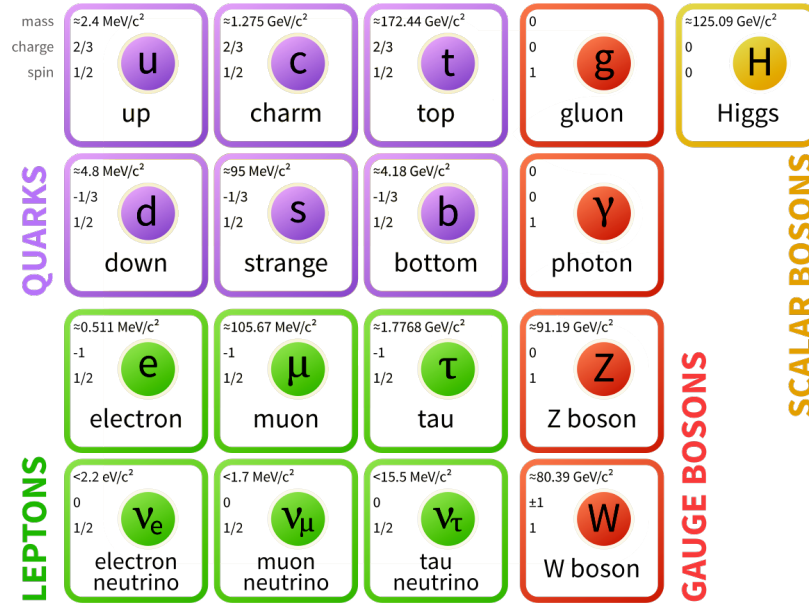


Figure 1.1: Particles in the Standard Model of particle physics [12].

gravity. Our hope with LHC searches is to find examples where the data doesn't "fit" the SM which will inform us how to extend the theory to Beyond the Standard Model (BSM), as illustrated in Figure 1.2. There are clues that the electroweak sector might provide the key for unlocking these mysteries. The confirmation of neutrino oscillations by the Super-Kamiokande experiment 25 years ago means that neutrinos have mass [13]. However, KATRIN [14] constrains the scale of neutrinos masses to be eleven orders smaller than other SM masses such as the top quark, and it's currently an open question what mechanism gives neutrinos their masses. Other anomalies from weak the weak sector include the 4.2σ tension from the recent $g - 2$ result [15] and the 7σ tension with the latest W mass measurement [16].

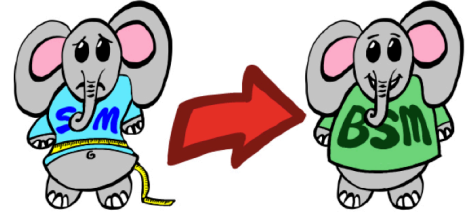


Figure 1.2: The goal for our physics searches: to find a model that better fits the data. Artist, Brendan Hartman.

An ansatz for the whole paradigm of electroweak symmetry breaking is the form of the Higgs potential, and the SM is the simplest solution that can give masses to the vector bosons. To test this formulation for the Higgs potential, we need to look for events with *two* Higgs bosons, or HH (read as "di-higgs") – and this is what this thesis is all about. This thesis takes advantage of the leading HH decay mode into four b -quarks. Although an advantage of this channel is the higher number of signal events to start with, it is complicated by a large background that is difficult to simulate from

first principles. But, as the size of our dataset grows, so too do the opportunities to come up with novel techniques that make $HH \rightarrow 4b$ increasingly more powerful. With the complex, multi-modal data that we have from our colliders, integrating recent developments from Machine Learning (ML) into physics analyses and object reconstruction is a hot-topic, and also a large part of this thesis work, both in terms of identifying the objects formed by b -quarks and for modeling the background processes.

This thesis is organized as follows. Part **I** reviews the introductory material for the thesis including the Standard Model (Chapter **2**), collider (Chapter **3**), experiment (Chapter **4**), and general event reconstruction (Chapter **5**). Part **II** explains the b -tagging algorithms, both for Run 2 (Chapter **6**) and with future looking optimizations (Chapter **7**). Part **III** details the $HH \rightarrow 4b$ analysis with the statistical procedure (Chapter **8**), analysis design (Chapter **10**), background estimation (Chapter **11**), and results. Finally, Part **IV** (Chapter **13**) introduces a novel background estimation procedure which could offer vital improvements for future iterations of this analysis. Chapter **14** summarizes with our conclusions.

Part I

Foundations

Here we review the introductory material for this thesis work. Chapter 2 introduces electroweak symmetry breaking and the Standard Model, and summarizes the experimental status of Higgs physics. Chapter 3 and Chapter 4 introduce the Large Hadron Collider and the ATLAS detector, respectively. Finally Chapter 5 describes how the electronic signals from the detector are reconstructed into physics objects.

2

The Standard Model

Surveying the rich experimental literature from which these examples are drawn makes one suspect that something like a paradigm is prerequisite to perception itself.

– Thomas Kuhn, *The Structure of Scientific Revolutions*

Particles and their interactions are characterized by group theory. A Lagrangian is specified which respects a set of symmetries, and the resulting physics minimizes the action as specified by the Lagrangian. The SM is the most general dimension 4 Lagrangian respecting the $U(1) \times SU(2) \times SU(3)$ symmetries. The $U(1)$ interactions correspond to electromagnetism (described in Chapter ??), while $U(1) \times SU(2)$ describe the unified electroweak interactions (described in Chapter 2.2), and $SU(3)$ describes QCD (described in Chapter 2.4). The SM Lagrangian in its most concise form is given by:

$$\mathcal{L}_{SM} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{\psi}\gamma_{\mu}D^{\mu}\psi + (y_{ij}\bar{\psi}_i\psi_j + h.c.) + |D^{\mu}\phi|^2 - V(\phi). \quad (2.1)$$

Feynman diagrams are the technology to the Lagrangian terms into a prediction. The Feynman terms for each of these Lagrangian terms are shown in Figure 2.1. The first term is the kinetic energy of the vector bosons. The second terms is the coupling the fermions to the corresponding gauge bosons. The third term is the Yukawa couplings (which will be described in Chapter 2.3.1 and Chapter 2.3.2) and the 4th is the Higgs interaction with the vector bosons. The last term is the Higgs potential – the part of \mathcal{L}_{SM} that is most crucial for this thesis.

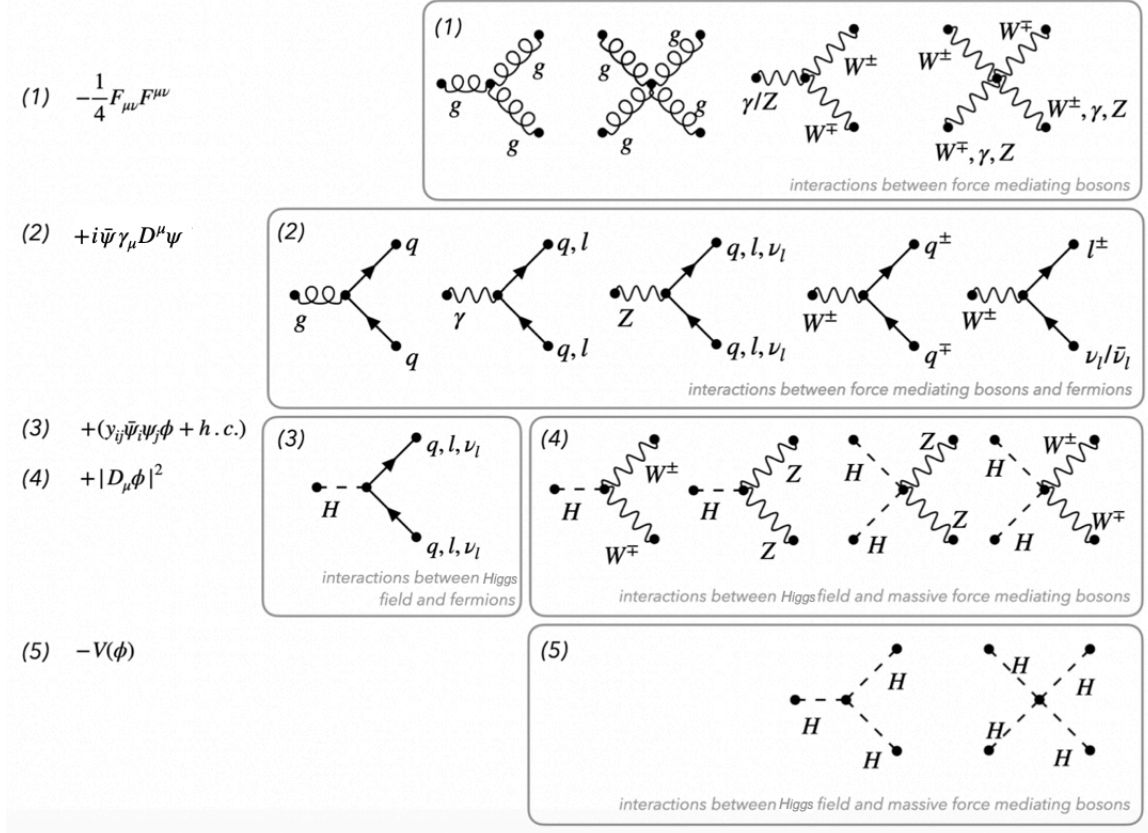


Figure 2.1: The Feynman diagrams corresponding to the terms in the Lagrangian [17].

2.1 Quantum Electro Dynamics

Photons

Photons are massless spin 0 particles. A non-interacting photon has a kinetic term with two derivatives:

$$F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu} \quad (2.2)$$

where A_{μ} is the four-vector potential from electromagnetism, which is also the photon field.

$$A_{\mu} = \epsilon_{\mu}(p)e^{-ip \cdot x} \quad (2.3)$$

The QED Lagrangian transforms under U(1) transformations:

$$A_{\mu} \rightarrow A_{\mu}(x) + Q\partial_{\mu}\alpha(x). \quad (2.4)$$

Spin 1/2 particles

The Dirac equation governing the dynamics of a spin 1/2 particle is

$$(i\gamma^\mu \partial_\mu - m)\Psi = 0, \quad \text{where } \gamma^\mu = \begin{pmatrix} 0 & \sigma_\mu \\ \bar{\sigma}_\mu & 0 \end{pmatrix}$$

, and

$$\sigma_\mu = (\mathbb{I}, \sigma_1, \sigma_2, \sigma_3), \quad \bar{\sigma}_\mu = (\mathbb{I}, -\sigma_1, -\sigma_2, -\sigma_3).$$

In the Weyl representation, the four component field Ψ gets broken down into two 2-component spinors $\Psi = \begin{pmatrix} \Psi_L \\ \Psi_R \end{pmatrix}$, where the L and R subscripts denote the spinors that are the left and right-handed spinors. So in the Weyl representation, the Dirac equation can be written as two separate equations,

$$i\sigma^\mu \partial_\mu \Psi_R - m\Psi_L = 0 \tag{2.5}$$

$$i\bar{\sigma}^\mu \partial_\mu \Psi_L - m\Psi_R = 0, \tag{2.6}$$

where the fermion mass term mixes the left and right-handed fields and couples these equations.

The corresponding Dirac Lagrangian for a free fermion is:

$$\mathcal{L}_D = i\Psi_R^\dagger \sigma^\mu \partial_\mu \Psi_R - m\Psi_R^\dagger \Psi_L + i\Psi_L^\dagger \bar{\sigma}^\mu \partial_\mu \Psi_L - m\Psi_L^\dagger \Psi_R \tag{2.7}$$

Then for interacting fermions with the photon field, the gauge invariance is preserved when we replace the derivatives with the covariant derivatives:

$$D_\mu = \partial_\mu + iQA_\mu. \tag{2.8}$$

2.2 The Higgs mechanism

The following section gratefully borrows from the pedagogical introductions in [18] and [9].

The weak force governs nuclear decay. In 1957, an avant-garde experiment by Chien-Shiung Wu demonstrated that parity was violated in weak interactions [19]. Additionally, the weak force *maximally* violates parity, and these interactions can be described by a “V-A” (or vector-axial) theory with the weak force’s charged current only coupling left-handed spinors. This V-A theory is an effective field theory with a four-point interaction term. The coupling of this four-point interaction term, is Fermi’s constant $G_F = 1.17 \cdot 10^{-5} \text{ GeV}^{-2}$. The smallness of this coupling is understood as being mediated by a massive particle with a coupling of the order: $m_W \sim \frac{1}{\sqrt{G_F}} = \mathcal{O}(100) \text{ GeV}$. In analogy with electromagnetism, the weak boson mediators should also be spin 1 particles.

However, a naive introduction of a mass term for a spin 1 particle violates the gauge invariance that we expect for physical solutions. The theory of superconductivity illustrates how a non-zero vacuum expectation value (vev)¹ can give rise to an effective mass for the photon through an interaction term with this vev. In direct analogy with the solution from superconductivity, we currently understand the weak bosons acquiring their mass by the Higgs mechanism.

Consider the symmetry group of $SU(2) \times U(1)$ transformations. This gives *four* gauge fields, which we will call: A^1, A^2, A^3 (for the gauge fields of the $SU(2)$ transformation) and B (corresponding to the $U(1)$ symmetry). An element of the $SU(2)$ group is specified by $e^{i\vec{\alpha} \cdot \frac{\vec{\sigma}}{2}}$, where $\alpha \in [0, 2\pi]$.

Consider a new field ϕ , which is a complex doublet that transforms under these $SU(2) \times U(1)$ transformations as:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \rightarrow e^{i\vec{\alpha} \cdot \vec{\sigma}/2} e^{i\beta/2} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \quad (2.9)$$

where this suggestive notation for ϕ^+ and ϕ^0 will become apparent soon.

Suppose that this new field has a potential given by:

$$V(\phi) = -\mu^2 |\phi|^2 + \lambda (|\phi|^2)^2 \quad (2.10)$$

Since each of the ϕ components is a complex field with a real and imaginary component, this potential $V(\phi)$ actually is evaluated in 4d, but a visualization of the form of the potential for one of the components is shown in Figure 2.2. The characteristic form of the potential is that the minimum value for the field does not correspond to the zero coordinate, and there is a manifold of solutions for non-zero coordinates that will minimize the potential.

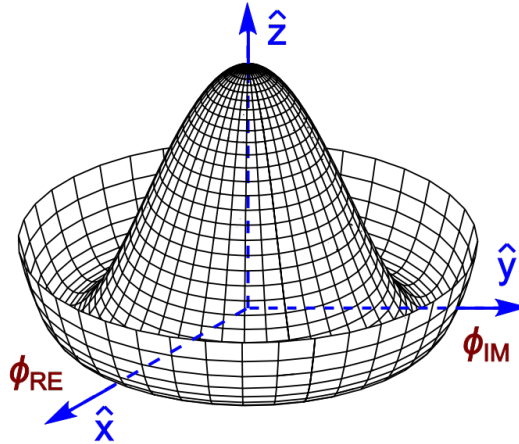


Figure 2.2: The form of the Higgs potential for one of the components [20].

¹A non-zero vacuum expectation value (vev) can apply the destruction operator to the ground state and get a non-zero solution.

The minimum of this potential is:

$$V'(\phi) = -2\mu^2\phi + 4\lambda\phi|\phi|^2 = 0 \quad (2.11)$$

$$\implies |\phi|^2 = \frac{1}{2} \frac{\mu^2}{\lambda}. \quad (2.12)$$

We define the “vacuum expectation value” v of the Higgs field as $v = \sqrt{2} \langle |\phi| \rangle = \frac{\mu}{\sqrt{\lambda}}$. We then reparametrize the field to expand around this minimum.

$$\phi(x) = \begin{pmatrix} \frac{1}{\sqrt{2}} \left(\pi^1(x) + i\pi^2(x) \right) \\ \frac{1}{\sqrt{2}} \left(v + h(x) + i\pi^3(x) \right) \end{pmatrix} \quad (2.13)$$

With the freedom of the 3 independent parameters specifying the parameters of the gauge field, we choose a gauge that transforms away three of the degrees of freedom for the field components of the ϕ vector. We choose to leave the real piece of the second component of the ϕ doublet as the final remaining degree of freedom:

$$\phi(x) = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} + \frac{h(x)}{\sqrt{2}} \end{pmatrix} \quad (2.14)$$

Now we associate the remaining $h(x)$ field as the *Higgs field*. We now insert this instantiation of $\phi(x)$ into the kinetic term of the Lagrangian to examine the coupling of the weak boson force fields with this vev.

The covariant derivative for the weak field is:

$$D_u \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = (\partial_\mu - igI^a A_\mu^a - ig'B_\mu Y) \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix} \quad (2.15)$$

where g is the coupling strength for the SU(2) subgroup, g' is the coupling strength of the U(1) subgroup, I^a are the weak isospin generators and Y is the U(1) hypercharge.

Next we plug in the generators of the SU(2) group: $I^a = \frac{1}{2}\sigma_a$ for $a = 1, 2, 3$, and $Y = \frac{1}{2}$ for the Higgs doublet.

$$D_u \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \left(-ig\frac{\sigma_a}{2} A_\mu^a - ig'\frac{B_\mu}{2} \right) \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix} + \dots \quad (2.16)$$

where the partial went go away from of the choice of setting the Higgs field to 0.

We simplify by plugging in the Pauli matrices:

$$g\frac{\sigma_a}{2} A_\mu^a + \frac{g'}{2} B_\mu = \begin{pmatrix} \frac{g}{2} A_\mu^3 + \frac{g'}{2} B_\mu & \frac{g}{2} (A_\mu^1 - iA_\mu^2) \\ \frac{g}{2} (A_\mu^1 + iA_\mu^2) & -\frac{g}{2} A_\mu^3 + \frac{g'}{2} B_\mu \end{pmatrix}, \quad (2.17)$$

then multiply in the Higgs field:

$$D_\mu \phi = -i \begin{pmatrix} \frac{g}{2} A_\mu^3 + \frac{g'}{2} B_\mu & \frac{g}{2} (A_\mu^1 - i A_\mu^2) \\ \frac{g}{2} (A_\mu^1 + i A_\mu^2) & -\frac{g}{2} A_\mu^3 + \frac{g'}{2} B_\mu \end{pmatrix} \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \frac{v}{\sqrt{2}} \frac{g}{2} (A_\mu^1 - i A_\mu^2) \\ \frac{v}{\sqrt{2}} (-\frac{g}{2} A_\mu^3 + \frac{g'}{2} B_\mu) \end{pmatrix} \quad (2.18)$$

The mass term from the interaction with the vev, will not depend on the coupling with the Higgs field, so we can set $h(x) \rightarrow 0$ in the following. So the kinetic term of the electroweak force in the Lagrangian is:

$$\mathcal{L}_K = |D_\mu \phi|^2 = \frac{v^2}{2} \frac{g^2}{4} (|A_\mu^1|^2 + |A_\mu^2|^2) + \frac{v^2}{2} \left(-\frac{g}{2} A_\mu^3 + \frac{g'}{2} B_\mu \right)^2. \quad (2.19)$$

Now we can identify the weak components of the weak isospin field with the charged W bosons:

$$W_\mu^+ \equiv \frac{A_\mu^1 - i A_\mu^2}{\sqrt{2}}, \quad W_\mu^- \equiv \frac{A_\mu^1 + i A_\mu^2}{\sqrt{2}}, \quad (2.20)$$

which together give the corresponding mass term in the Lagrangian:

$$\mathcal{L}_K \supset \left(\frac{vg}{2} \right)^2 W_\mu^+ W^{\mu,-}. \quad (2.21)$$

This means the W mass is $\boxed{m_W = \frac{gv}{2}}$.²

To express this field mixing succinctly, let $\tan \theta_W \equiv g'/g$, so that $\cos \theta_W = \frac{g}{\sqrt{g'^2 + g^2}}$ and $\sin \theta_W = \frac{g'}{\sqrt{g'^2 + g^2}}$. Then we can associate the last term mixing the two fields with the last massive boson.

$$Z_\mu = \cos \theta_W A_\mu^3 - \sin \theta_W B_\mu, \quad (2.22)$$

which has a mass of $\boxed{m_Z = \frac{1}{2} v \sqrt{g^2 + g'^2}}$

We can find the linear combination of the fields that corresponds to the massless photon by orthogonality with the Z_μ field.

$$A_\mu = \sin \theta_W A_\mu^3 + \cos \theta_W B_\mu \quad (2.23)$$

Then we can invert these equations to write down the covariant derivative in terms of the W , Z and γ fields of the broken theory.

²Normally there is a factor of $\frac{1}{2}$ in the mass term of the Lagrangian that we associate with the mass, i.e, for the Klein Gordon equation $\mathcal{L} = \frac{1}{2} \partial^\mu \partial_\mu \phi - \frac{1}{2} m^2 \phi^2$, and the mass of the field is ϕ . However, since this term is giving a mass to the W^+ boson and it's antiparticle, this why this coupling is twice as large as if this was the term for a field that was also its own anti-particle (such as the Z -boson or the Higgs boson).

$$\begin{aligned}
A_\mu^1 &= \frac{W_\mu^+ + W_\mu^-}{\sqrt{2}}, & A_\mu^2 &= \frac{iW_\mu^+ - iW_\mu^-}{\sqrt{2}} \\
A_\mu^3 &= \cos \theta_W Z_\mu + \sin \theta_W A_\mu \\
B_\mu &= -\sin \theta_W Z_\mu + \cos \theta_W A_\mu
\end{aligned} \tag{2.24}$$

Inserting these into $D_\mu = \partial_\mu - i\frac{g}{2}(\sigma_1 A_\mu^1 + \sigma_2 A_\mu^2 + \sigma_3 A_\mu^3) - ig'\frac{Y}{2}B_\mu$ and grouping together the physical fields gives:

$$\begin{aligned}
D_\mu &= \partial_\mu - i\frac{g}{\sqrt{2}}\frac{\sigma_1 + i\sigma_2}{2}W_\mu^+ - i\frac{g}{\sqrt{2}}\frac{\sigma_1 - i\sigma_2}{2}W_\mu^- \\
&\quad - i\left(g\cos\theta_W\frac{\sigma_3}{2} - g'\sin\theta_W Y\frac{\mathbb{I}}{2}\right)Z_\mu - i\left(g\sin\theta_W\frac{\sigma_3}{2} + g'\cos\theta_W Y\frac{\mathbb{I}}{2}\right)A_\mu.
\end{aligned} \tag{2.25}$$

And finally, we substitute the g and g' couplings back in for θ_W .

$$\begin{aligned}
D_\mu &= \partial_\mu - i\frac{g}{\sqrt{2}}\frac{\sigma_1 + i\sigma_2}{2}W_\mu^+ - i\frac{g}{\sqrt{2}}\frac{\sigma_1 - i\sigma_2}{2}W_\mu^- \\
&\quad - i\left(\frac{g^2}{\sqrt{g^2 + g'^2}}\frac{\sigma_3}{2} - \frac{g'^2}{\sqrt{g^2 + g'^2}}Y\frac{\mathbb{I}}{2}\right)Z_\mu - i\left(\frac{\sigma_3 + Y\mathbb{I}}{2}\right)\frac{gg'}{\sqrt{g^2 + g'^2}}A_\mu
\end{aligned} \tag{2.26}$$

Since A_μ is the photon, we can see that $e = \frac{gg'}{\sqrt{g^2 + g'^2}}$, is the fundamental unit of charge from

EM. For the Higgs doublet $Y = 1$, so $\frac{\sigma_3 + \mathbb{I}}{2} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ gives the electric charge of each component of the 2 Higgs doublet model. The vev v can be written down in terms of G_F , which is measured from the muon lifetime experiments.

Although the symmetry of the underlying solutions is no longer manifest, the structure of the underlying symmetries makes this a highly predictive theory. Now we know *almost* everything. The last step is knowing m_H , which we don't know without λ a priori:

$$m_H = \sqrt{2\lambda}v. \tag{2.27}$$

We'll revise the Higgs mass discovery and studies of the Higgs sector in Chapter 2.5. But first, we'll go through the rest of the SM, starting off with how the Higgs doublet also gives mass to the fermions (quarks and charged leptons).

2.3 Electroweak interactions with Fermions

2.3.1 Leptons

Yukawa interactions

With the Higgs doublet, we additionally will have the coupling between the fermions and gauge bosons via this covariant derivative.

$$\mathcal{L} \supset -y_e \begin{pmatrix} \nu_{eL}^\dagger & e_L^\dagger \end{pmatrix} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} e_R - y_\mu \begin{pmatrix} \nu_{\mu L}^\dagger & \mu_L^\dagger \end{pmatrix} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \mu_R - y_\tau \begin{pmatrix} \nu_{\tau L}^\dagger & \tau_L^\dagger \end{pmatrix} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tau_R + \text{h.c.} \quad (2.28)$$

We set ϕ to its vev and simplify:

$$\begin{aligned} \mathcal{L} &\supset -y_e \begin{pmatrix} \nu_{eL}^\dagger & e_L^\dagger \end{pmatrix} \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix} e_R - y_\mu \begin{pmatrix} \nu_{\mu L}^\dagger & \mu_L^\dagger \end{pmatrix} \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix} \mu_R - y_\tau \begin{pmatrix} \nu_{\tau L}^\dagger & \tau_L^\dagger \end{pmatrix} \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix} \tau_R + \text{h.c.} \\ &= \frac{y_e v}{\sqrt{2}} e_L^\dagger e_R + \frac{y_\mu v}{\sqrt{2}} \mu_L^\dagger \mu_R + \frac{y_\tau v}{\sqrt{2}} \tau_L^\dagger \tau_R + \text{h.c.} \end{aligned} \quad (2.29)$$

The vev is only couples to the lower component of the left-handed doublet, and this gives the charged leptons masses, while leaving the neutrinos massless. This description is valid for the physics analyses at the Large Hadron Collider, so we'll leave the neutrinos as massless for the rest of this presentation.³

We can write these terms more succinctly with a diagonal matrix:

$$\mathcal{L} \supset \begin{pmatrix} e_L^\dagger & \mu_L^\dagger & \tau_L^\dagger \end{pmatrix} \begin{pmatrix} \frac{y_e v}{\sqrt{2}} & & \\ & \frac{y_\mu v}{\sqrt{2}} & \\ & & \frac{y_\tau v}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} e_R \\ \mu_R \\ \tau_R \end{pmatrix} + \text{h.c.} \quad (2.30)$$

This diagonal matrix has the Yukawa couplings that give the leptons their mass $m_\ell = \frac{y_\ell v}{\sqrt{2}}$

Weak interactions

To find the interaction with the weak bosons, we use the term $i\bar{\psi}\gamma_\mu D^\mu\psi$, where the ψ are in the Weyl representation. We start from Eq. 2.7 and replace the ∂_μ with the covariant derivative D_μ to preserve gauge invariance. Below, we just focus on the terms of the W-boson weak interactions with the weak isodoublet, so the terms of $D_\mu = \partial_\mu - i\frac{g}{\sqrt{2}}\left(\frac{\sigma_1 + i\sigma_2}{2}W_\mu^+ + \frac{\sigma_1 - i\sigma_2}{2}W_\mu^-\right) + \dots$

³Neutrino oscillations means that (at least two) neutrinos do have a non-zero mass. Including neutrinos would give another flavor / mass unitary mixing matrix (the PMNS matrix), adding 3 angles and one complex phase to our accounting of the number of SM parameters.

$$\mathcal{L} \supset \begin{pmatrix} \nu_{eL}^\dagger & e_L^\dagger \end{pmatrix} i\bar{\sigma}^\mu \left(-i\frac{g}{\sqrt{2}} \right) \left(\frac{\sigma_1 + i\sigma_2}{2} W_\mu^+ + \frac{\sigma_1 - i\sigma_2}{2} W_\mu^- \right) \begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix} \quad (2.31)$$

We can simplify the Pauli matrices:

$$\frac{\sigma_1 + i\sigma_2}{2} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \frac{\sigma_1 - i\sigma_2}{2} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \quad (2.32)$$

$$\mathcal{L} \supset \begin{pmatrix} \nu_{eL}^\dagger & e_L^\dagger \end{pmatrix} i\bar{\sigma}^\mu \left(-i\frac{g}{\sqrt{2}} \right) \begin{pmatrix} 0 & W_\mu^+ \\ W_\mu^- & 0 \end{pmatrix} \begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix} = \frac{g}{\sqrt{2}} \nu_{eL}^\dagger \bar{\sigma}^\mu W_\mu^+ e_L + \frac{g}{\sqrt{2}} e_L^\dagger \bar{\sigma}^\mu W_\mu^- \nu_{eL} \quad (2.33)$$

Note - each of these terms individually has a neutral electric charge since each of these interactions respect charge conservation. Since there are also terms for the rest of the lepton generations, all six of these terms can be expressed succinctly with an inner product over the generations:

$$\mathcal{L} \supset \frac{g}{\sqrt{2}} \begin{pmatrix} \nu_{eL}^\dagger & \nu_{\mu L}^\dagger & \nu_{\tau L}^\dagger \end{pmatrix} \bar{\sigma}^\mu W_\mu^+ \begin{pmatrix} e_L \\ \mu_L \\ \tau_L \end{pmatrix} + \text{h.c.} \quad (2.34)$$

The above took the same flavor basis (for the interactions with the gauge bosons) as the mass eigenbasis. In general (as we'll see next for the quarks) these two bases are *not* simultaneously diagonalizable. However, when the neutrino masses are zero, we are free to choose any unitary matrix to mix the flavor states of the neutrinos, which we choose such that the flavor and mass matrices are simultaneously diagonalizable – justifying the above presentation.

2.3.2 Quarks

Yukawa interactions

Both the up and down type quarks are massive, so we need the vev to interact with both components of this field. Since the W^+ takes a positive unit of charge to turn a positron into a neutrino (or an up-type quark into a down-type quark), by analogy, we can see that we will need a W^- boson to turn a down-type quark into an up-type quark, which guides how to set up the Lagrangian for the quark Yukawa terms:

$$\mathcal{L} \subset -y_d \begin{pmatrix} u_L^\dagger & d_L^\dagger \end{pmatrix} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} d_R - y_s \begin{pmatrix} c_L^\dagger & s_L^\dagger \end{pmatrix} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} s_R - y_b \begin{pmatrix} t_L^\dagger & b_L^\dagger \end{pmatrix} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} b_R \quad (2.35)$$

$$- y_u \begin{pmatrix} u_L^\dagger & d_L^\dagger \end{pmatrix} \begin{pmatrix} \phi^0 \\ \phi^- \end{pmatrix} u_R - y_c \begin{pmatrix} c_L^\dagger & s_L^\dagger \end{pmatrix} \begin{pmatrix} \phi^0 \\ \phi^- \end{pmatrix} c_R - y_t \begin{pmatrix} t_L^\dagger & b_L^\dagger \end{pmatrix} \begin{pmatrix} \phi^0 \\ \phi^- \end{pmatrix} t_R + \text{h.c.} \quad (2.36)$$

We can still expand about the vev for $\phi^0 = \frac{v}{\sqrt{2}}$, $\phi^+ = \phi^- = 0$. This gives:

$$\mathcal{L} \subset -\frac{y_d v}{\sqrt{2}} d_L^\dagger d_R - \frac{y_s v}{\sqrt{2}} s_L^\dagger s_R - \frac{y_b v}{\sqrt{2}} b_L^\dagger b_R - \frac{y_u v}{\sqrt{2}} u_L^\dagger u_R - \frac{y_c v}{\sqrt{2}} c_L^\dagger c_R - \frac{y_t v}{\sqrt{2}} t_L^\dagger t_R + \text{h.c.} \quad (2.37)$$

or in the mass diagonal basis:

$$\mathcal{L} \supset -\begin{pmatrix} d_L^\dagger & s_L^\dagger & b_L^\dagger \end{pmatrix} \begin{pmatrix} \frac{y_d v}{\sqrt{2}} & & \\ & \frac{y_s v}{\sqrt{2}} & \\ & & \frac{y_b v}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} d_R \\ s_R \\ b_R \end{pmatrix} - \begin{pmatrix} u_L^\dagger & c_L^\dagger & t_L^\dagger \end{pmatrix} \begin{pmatrix} \frac{y_u v}{\sqrt{2}} & & \\ & \frac{y_c v}{\sqrt{2}} & \\ & & \frac{y_t v}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} u_R \\ c_R \\ t_R \end{pmatrix} + \text{h.c.} \quad (2.38)$$

Again, we have the fermions getting their masses from interaction with the Higgs vev, and $m_q = \frac{y_q v}{\sqrt{2}}$.

Weak interactions

For the weak interactions, now we will denote the quark flavor eigenstates with an “ F ” subscript. Similar to what we saw for the leptons, the interactions with the W bosons can be expressed as:

$$\mathcal{L} \supset \frac{g}{\sqrt{2}} \begin{pmatrix} u_{L,F}^\dagger & c_{L,F}^\dagger & t_{L,F}^\dagger \end{pmatrix} \bar{\sigma}^\mu W_\mu^+ \begin{pmatrix} d_{L,F} \\ s_{L,F} \\ b_{L,F} \end{pmatrix} + \text{h.c.} \quad (2.39)$$

The mass and flavor bases for the up and down type quarks can be related by a unitary transformation:

$$\begin{pmatrix} d_{L,F} \\ s_{L,F} \\ b_{L,F} \end{pmatrix} = \mathbb{U}_d \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix}, \quad \begin{pmatrix} u_{L,F} \\ c_{L,F} \\ t_{L,F} \end{pmatrix}^\dagger = \mathbb{U}_u^\dagger \begin{pmatrix} u_L \\ c_L \\ t_L \end{pmatrix}^\dagger. \quad (2.40)$$

We can plug these expressions back into the Lagrangian to get:

$$\mathcal{L} \supset \frac{g}{\sqrt{2}} \begin{pmatrix} u_L^\dagger & c_L^\dagger & t_L^\dagger \end{pmatrix} V_{\text{CKM}} \bar{\sigma}^\mu W_\mu^+ \begin{pmatrix} d_L \\ s_L \\ b_L \end{pmatrix} + \text{h.c.}, \quad (2.41)$$

where $V_{\text{CKM}} = \mathbb{U}_u^\dagger \mathbb{U}_d$ is the Cabibbo-Kobayashi-Maskawa (CKM) matrix. Since both generations of quarks are massive, this matrix is non-diagonalizable and dictates the degree to which the flavor interactions couple the quarks between each other. This V_{CKM} is a 3×3 unitary matrix, which in general would need to be represented by 9 numbers, the three Euler angles, and 6 complex phases. However, five of these phases can be absorbed by redefining the phases on the quark fermion fields. This still leaves one phase left over since the Lagrangian needs to stay real.

The common parametrization of the CKM matrix is:

$$V_{\text{CKM}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_{23} & \sin \theta_{23} \\ 0 & -\sin \theta_{23} & \cos \theta_{23} \end{pmatrix} \begin{pmatrix} \cos \theta_{13} & 0 & e^{-i\delta} \sin \theta_{13} \\ 0 & 1 & 0 \\ -e^{i\delta} \sin \theta_{13} & 0 & \cos \theta_{13} \end{pmatrix} \begin{pmatrix} \cos \theta_{12} & \sin \theta_c & 0 \\ -\sin \theta_c & \cos \theta_c & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (2.42)$$

These angles and phases are measured to be: $\sin \theta_{12} = 0.2245$, $\sin \theta_{23} = 0.042$, $\sin \theta_{13}$, and $\delta = 1.2$. The left-over phase δ causes the particle and anti-particle interactions to be *different*, resulting in CP violation in the quark sector. The size of the CP violation term is controlled by $\sin \theta_{13}$, so since $\sin \theta_{13}$ is small, the CP violation in the quark sector isn't large. When these angles and phases are plugged in and the matrices are multiplied out, the CKM matrix is:

$$V_{\text{CKM}} = \begin{pmatrix} 0.972 & 0.225 & 0.003 \\ 0.225 & 0.973 & 0.041 \\ 0.009 & 0.040 & 0.999 \end{pmatrix} \quad (2.43)$$

It's nearly diagonal, which means that the quarks couple most strongly within a generation.

2.4 Quantum Chromodynamics

There is one last force for the interactions of quarks, which is Quantum Chromodynamics (QCD), which has some interesting phenomenology for collider events.. The mediator of the strong force is the *gluon*, a massless, spin 1 particle:

$$A_\mu^a(x) \rightarrow A_\mu^a(x) + \frac{1}{g_s} \partial_\mu \alpha^a(x) - f^{abc} \alpha^b(x) A_\mu^c(x), \quad (2.44)$$

where g_s is the coupling of QCD, also often called α_s . QCD in some sense is similar to QED, except it is a *non-Abelian* gauge theory (the $f^{abc} \neq 0$ in QCD), which gives gluon self-interaction vertices, as shown in Figure 2.1. Just as the generator for fermions (in the SU(2) group) are the Pauli spinors,

the generators for the QCD group are given by the eight Gell-Mann matrices:

$$\begin{aligned} \lambda^1 &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ & & 0 \end{pmatrix}, \quad \lambda^2 = \begin{pmatrix} 0 & -i \\ i & 0 \\ & & 0 \end{pmatrix}, \quad \lambda^3 = \begin{pmatrix} 1 & & \\ & -1 & \\ & & 0 \end{pmatrix}, \quad \lambda^4 = \begin{pmatrix} & & 1 \\ & 0 & \\ 1 & & \end{pmatrix}, \\ \lambda^5 &= \begin{pmatrix} 0 & -i \\ & 0 & \\ i & & 0 \end{pmatrix}, \quad \lambda^6 = \begin{pmatrix} 0 & & \\ & 0 & 1 \\ & 1 & 0 \end{pmatrix}, \quad \lambda^7 = \begin{pmatrix} 0 & & \\ & 0 & -i \\ & i & 0 \end{pmatrix}, \quad \lambda^8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & & \\ & 1 & \\ & & -2 \end{pmatrix}. \end{aligned} \quad (2.45)$$

The group describing the strong force is the SU(3) group, and the unit of charge that it carries is the *color charge*. Although many experiments tried to isolate a quark, they all proved unsuccessful. Each of the quarks carries a *color*, which could be red, green, blue.⁴ We now understand the result of not seeing any free quarks in nature by a postulate, the *confinement principle*, that says there are no free objects with color. In analogy with light, the combination of all three of these colors gives a colorless particle. Experimentally, this means we observe either a baryon with three quarks carrying the red, green, and blue color or a meson, a quark anti-quark pair carrying the color and corresponding anti-color. A *hadron* is either a meson or a baryon.

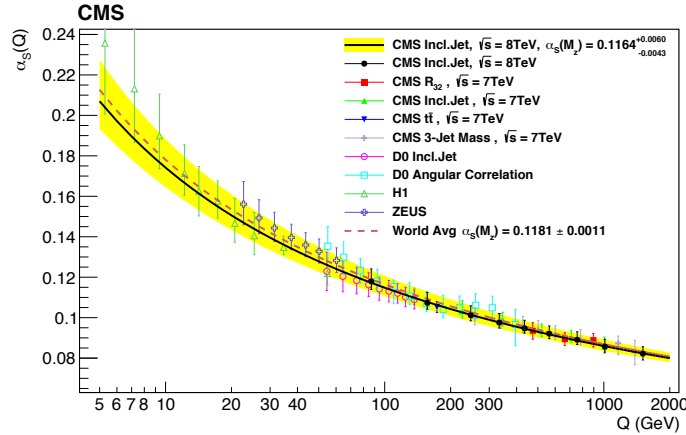


Figure 2.3: Collection of measurements illustrating the running of α_S [21]. The strength of the coupling decreases as the energy, Q , increases.

Figure 2.3 shows how the strength of the coupling decreases as the energy scale increases. So the mechanics of perturbation theory will only be valid for high energies. This also means that the force increases with distance (exactly opposite to the the $\frac{1}{r^2}$ force field from gravity and electromagnetism).

Two aspects in which QCD is important for this thesis are: (1) how quarks are observed experimentally as *jets* and (2) how we search for new physics with the collisions of the constituents of

⁴Calling this property color is purely whimsical as it doesn't have to do with the colors from the reflectivity of light.

protons, which are described by the *parton distribution functions*. These topics are briefly explained in the following subsections.

2.4.1 Jets

Figure 2.4(a) illustrates what happens to a quark as it gets pulled away from another quark. The force between the two quarks increases with the distance, until it becomes energetically favorable for quark / anti-quark pairs to pop out of the vacuum to balance the color charge imbalance by forming color-neutral hadrons.

For the quarks produced by electroweak interactions at the LHC, the energies are high and GeV range. The strength of the coupling at the Z mass is $\alpha_S(m_Z) = 0.12$, so the mechanics of perturbation theory are valid, and we can treat the quarks and gluons like useful quasiparticles [22]. Additionally, at high energies, QCD is a conformal field theory, which means the force doesn't have an intrinsic scale. (Strictly this is not true, because Figure 2.3 shows the running of the coupling with the energy scale, but at high energies the slope of α_S is close enough to horizontal that this is a useful approximation to understand the jet formation.)

A quark flying away from the collision point will radiate a gluon and then the gluon will pair produce a quark / anti-quark pair.⁵ This process continues in a cascade reaction until the energy of the final quarks and gluons is low enough (below a cutoff scale of $\Lambda_{QCD} \approx 100$ GeV) that they will form stable hadrons. These final hadrons give a collimated stream of particles going in the direction of original quark, as shown in Figure 2.4(b). A clustering algorithm defines the jet, which is a proxy for the kinematics of the original quark, and Chapter 5.3 describes how these jets are defined.

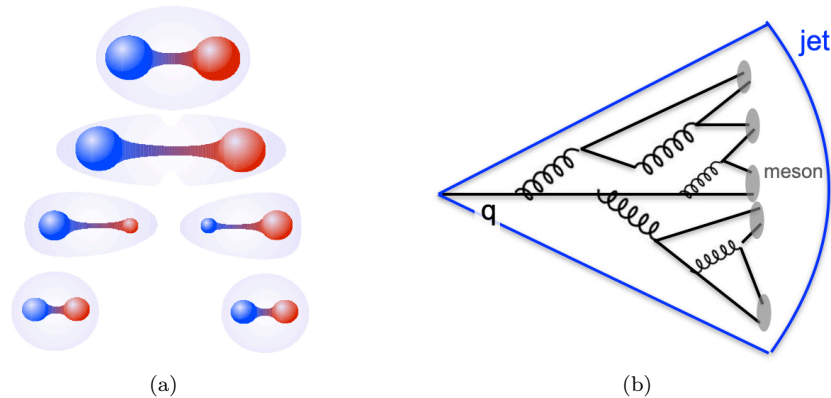


Figure 2.4: **Left:** Confinement principle illustration [24]. As two quarks are pulled apart, there will be a new quark / anti-quark pair to keep each object color neutral. **Right:** Illustration of the formation of a jet.

⁵A gluon also creates a similar signature in the detector, but the first step has the gluon splitting into a $q\bar{q}$ pair. The discrimination of jets initiated by quarks and gluons is an active area of research in collider physics [23].

2.4.2 Parton distribution functions

This thesis works with a dataset created from proton collisions. A proton is made up of two up quarks (charge $+\frac{2}{3}$) and one down quark (charge $-\frac{1}{3}$), which are called “valence quarks” and dictate the quantum numbers of the proton (such as charge and spin). The strong force holds the proton together via the gluons. In the midst of these gluons, there are also numerous “sea quarks” which can pop into existence out of the vacuum, as shown in Figure 2.5.

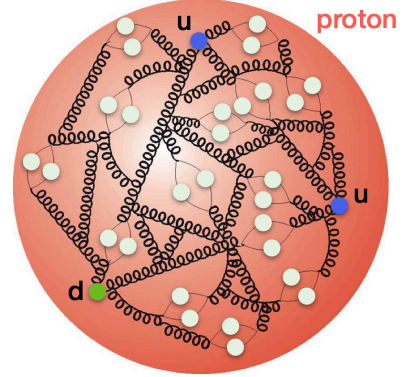


Figure 2.5: Partons in the proton, with the valence quarks in color, and the sea quarks in white [25].

Each parton, quark and gluon constituent i in the proton, will only carry a fraction x_i of the proton’s momentum. Summing over the partons in each proton gives the total pp cross-section:

$$\sigma_{A,B \rightarrow X} = \sum_{a,b} \int_0^1 dx_1 dx_2 f_a^A(x_1) f_b^B(x_2) \sigma_{a,b \rightarrow X}, \quad (2.46)$$

where A, B are the protons, a and b are the constituents, and the functions $f_a^A(x_1)$ and $f_b^B(x_2)$ are the parton distribution functions (pdfs). These pdfs depend on the momentum transfer, Q , of the interaction, as shown in Figure 2.6. At $Q^2 = 10 \text{ GeV}^2$ the peaks for the valence (up and down) quark pdfs at high x_i are clear. At high x , $f_u(x)$ is approximately twice as large as $f_d(x)$. Higher energies probe shorter distances, and at $Q^2 = 10^4 \text{ GeV}^2$ the sea quarks and gluons start to carry more of the proton’s momentum.

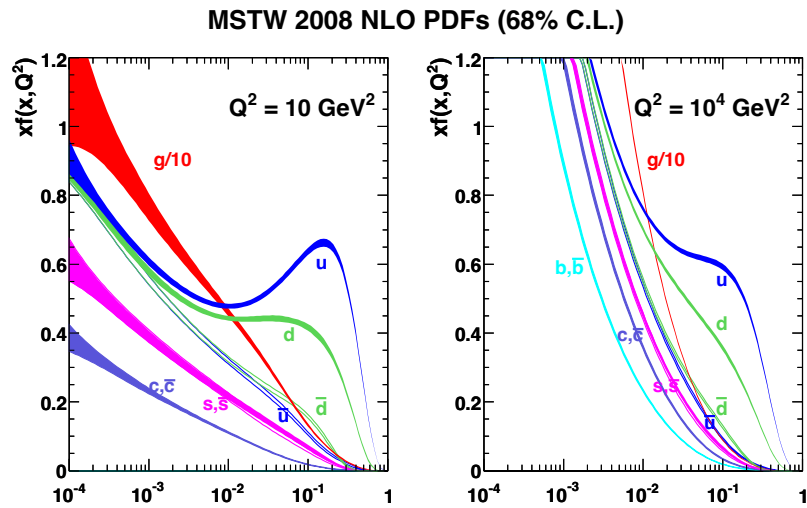


Figure 2.6: The proton’s pdfs for $Q^2 = 10 \text{ GeV}^2$ (left) and $Q^2 = 10^4 \text{ GeV}^2$ (right) [26].

The partons carry a range of momenta, making a proton collider an ideal discovery machine as a range of energies can be probed with one machine setting.

Figure 2.7 summarizes the steps described in the previous two subsections. The partons collide and the electroweak $t\bar{t}H$ event is shown in the red blob in the top hemisphere of the event. Also visible at this hadron collider are the high number of jets that are produced, making these complex events to reconstruct and analyze, which will be further described in Chapter 5.

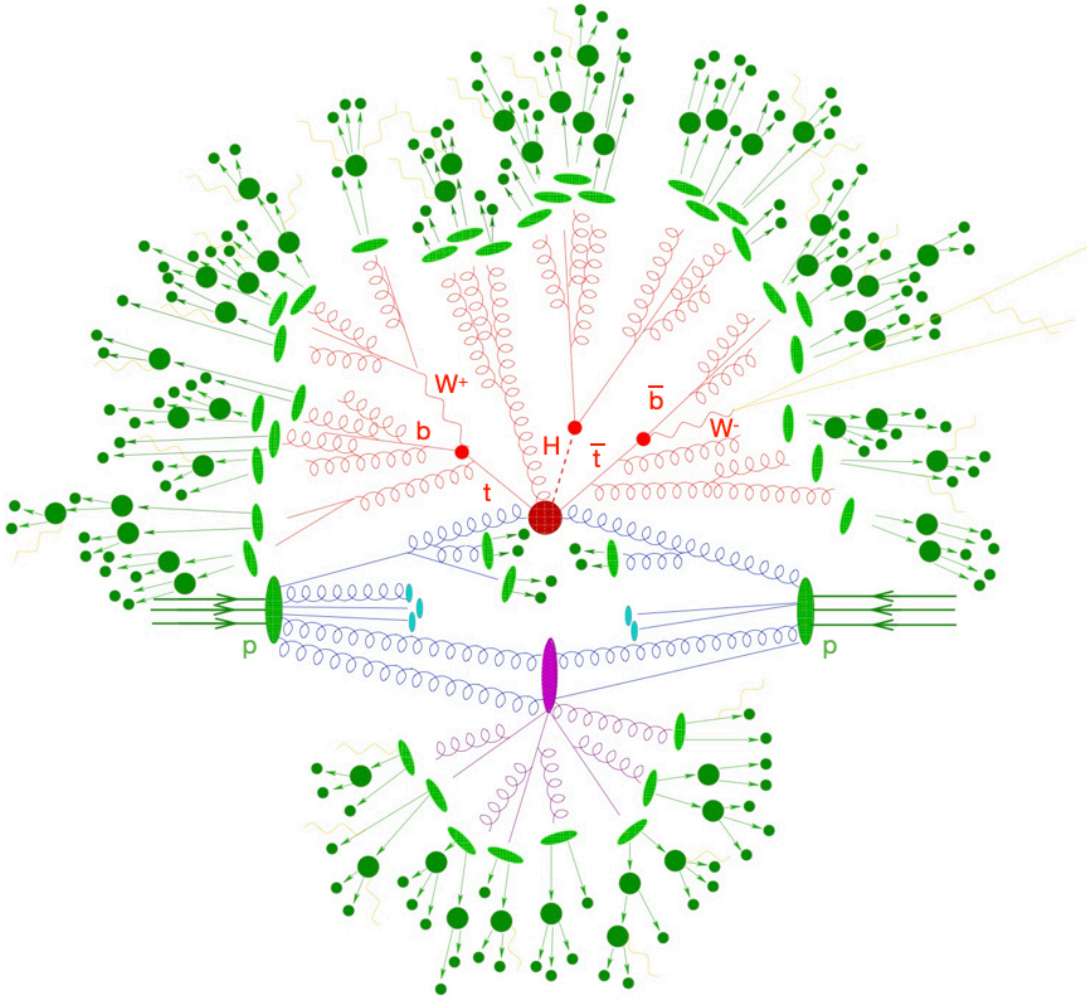


Figure 2.7: Event from a Monte Carlo generator [27]. The protons come in from the left and right, and the top red circle shows two gluons giving a $t\bar{t}H$ event. The lower purple blob is another parton collision coming from the same protons. The quarks and gluons fragment into other quarks and gluons. The light green blobs show the circles condensing into hadrons, while the dark green circles show the hadron decays.

2.5 What we've learned about the Higgs

The final free parameter of the SM to be measured was the Higgs mass, but a priori, there was little theoretical guidance for what mass scale it might be at. This is underscored by a quote from John Ellis, *et. al.* back in 1975:

We should perhaps finish our paper with an apology and a caution. We apologize to experimentalists for having no idea what is the mass of the Higgs boson, ..., and for not being sure of its couplings to other particles, except that they are probably all very small. [28]

Figure 2.8(a) indicates how the Higgs mass impacts the modes of particles it can decay into. Over the past half-century, the experiments started chipping away at the allowed parameter space. In the e^+e^- collisions at LEP, searches from the Higgs produced in association with the Z boson gave a lower bound of the Higgs mass of 114.5 GeV. Searching for the Higgs in $H \rightarrow WW$ decays at the Tevatron (and early ATLAS and CMS LHC results) gave upper limits on the Higgs mass around 130 GeV [18]. Additionally, fits to the electroweak data (including the W mass and electroweak precision observables) as a function of the Higgs mass indicated the plausible range for m_H , and these constraints from the beginning of 2012 are summarized in Figure 2.8(b). It was becoming increasingly clear that there was a narrow allowed mass range for the SM Higgs.

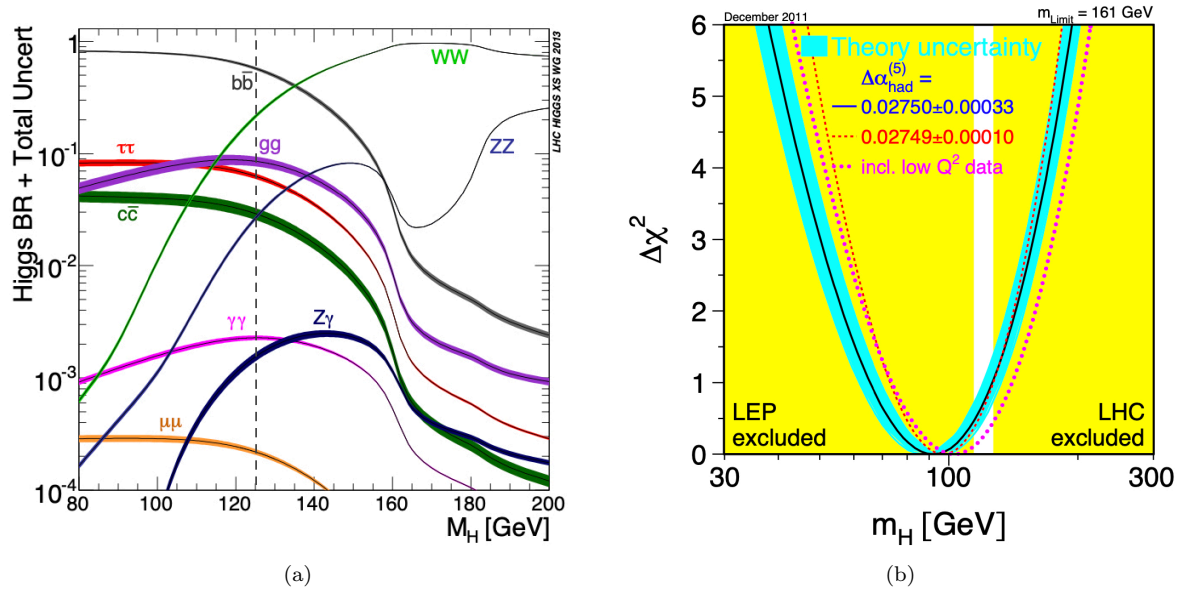


Figure 2.8: **Left:** The Higgs branching ratio for the leading decay modes as a function of the Higgs mass [29]. The dashed line indicates the measured Higgs mass at 125 GeV. **Right:** Allowed mass range for the Higgs boson at the beginning of 2012 [28]. The yellow regions are excluded by LEP and the Tevatron and LHC, while the parabolas indicate the $\Delta\chi^2 = \chi^2 - \chi^2_{\text{min}}$ from the electroweak fits.

On July 4th 2012 that the ATLAS and CMS experiments officially announced their discovery of the Higgs boson with a mass of 125 GeV [10, 11]. These discoveries were made in the clean decay modes with $H \rightarrow ZZ^* \rightarrow 4\ell$ and $H \rightarrow \gamma\gamma$, and updated versions of these discovery plots with the latest dataset from ATLAS are shown in Figure 2.9.

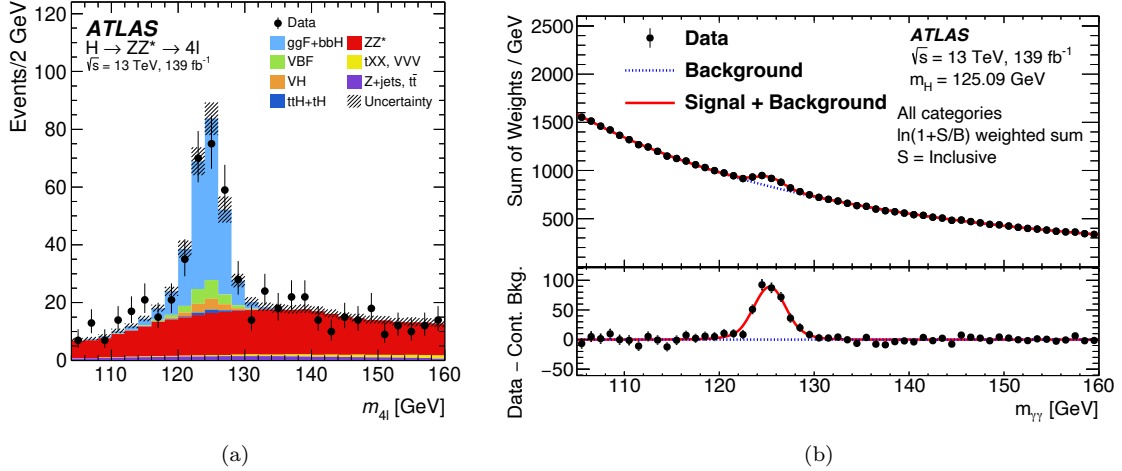


Figure 2.9: The invariant mass in the two channels used for the Higgs discovery: $H \rightarrow ZZ \rightarrow 4\ell$ [30] (left) and $H \rightarrow \gamma\gamma$ [31] (right).

After determining the Higgs mass, the effort began to test the sundry SM predictions to see if the new particle was the SM Higgs boson, or indicative of BSM physics. One such test is the Higgs production modes, which are shown in Figure 2.10.

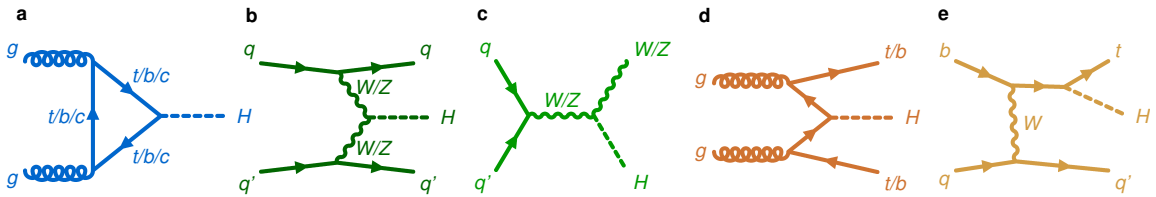


Figure 2.10: Higgs production modes at the LHC: (a) ggF, (b) VBF, (c) VH, (d) ttH, and (e) tH [32].

The leading production mode is gluon-gluon fusion (ggF) accounting for 87% of the overall cross-section. The massless gluon has no direct coupling with the Higgs, but this production is dominant because of the large top Yukawa coupling in the loop, and also because of the large gluon pdfs (as shown in Figure 2.6). The next leading production modes involve couplings to the electroweak bosons with vector boson fusion (VBF) accounting for 7% and higgsstrahlung, with a Higgs radiating off a W or Z vector boson (VH), accounting for 4%. The last production modes take advantage of the large top Yukawa coupling with ttH and tH accounting for 1% and 0.05% of the cross-section,

respectively.⁶ In the 10 years since the Higgs discovery, Figure 2.11(a) shows we have now observed the four leading production modes consistent with the SM expectation, and the observed (expected) upper limit on $t\bar{t}H$ production is 15 (7) times the SM prediction.

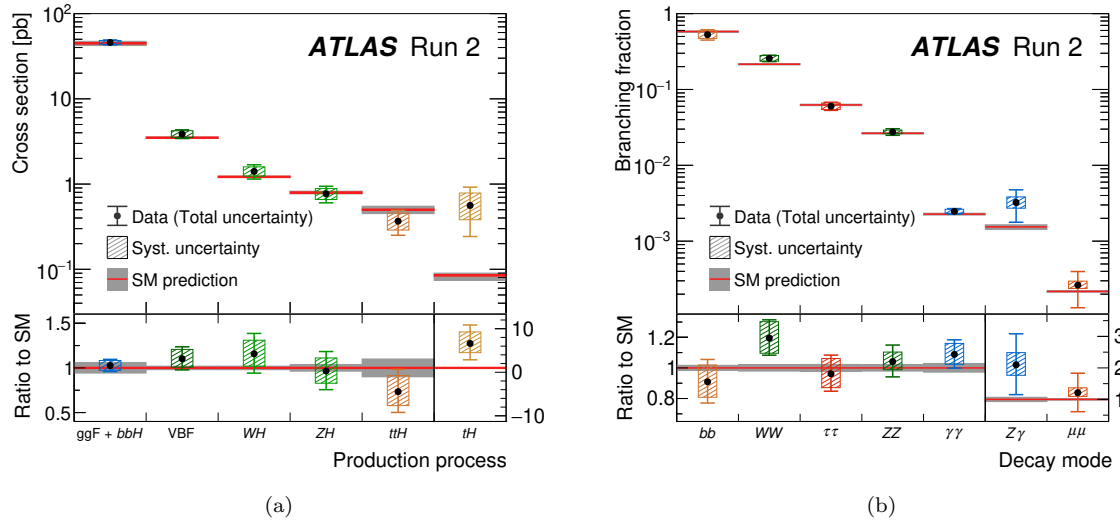


Figure 2.11: The Higgs production (left) and decay (right) modes. In the subpanels, the vertical line delineates the processes with a 5σ observation and the process that have lower significance [32].

Furthermore, Figure 2.8(a) shows that a 125 GeV Higgs mass fortuitously opens up a large number of decay modes with non-negligible branching ratios to test the Higgs sector. In the past 10 years, we've started searching for almost all of these decay modes shown. Figure 2.11(b) shows we've observed the decays to the kinematically accessible 3rd generation fermions, as well as the electroweak bosons. Excitingly, in Run 2 we've also started to gain sensitivity to the rarer modes, with $H \rightarrow \mu\mu$ having an observed (expected) significance of 2.0σ (1.7σ), and $H \rightarrow Z\gamma$ having an observed (expected) significance of 2.3σ (1.1σ).

Additionally, as described earlier in this chapter, the SM predicts that the fermion couplings scale linearly with the masses, and with the square of the vector boson masses. Figure 2.12 shows good agreement with the SM prediction. However, the log-log on plot does mean that the errors on the couplings to 2nd generation particles are quite a bit larger, and currently, there are observed (expected) upper limits for the charm coupling of 5.7 (7.6) times less than the SM value.

⁶The corresponding associated production with b -quarks, $b\bar{b}H$, is experimentally indistinguishable from ggF as there is a lower efficiency for identifying the b -jets in the forward region of the detector. The $b\bar{b}H$ production mode is added to ggF in the signal strength fit (adding 1% to the ggF cross-section) and these signal strengths are fit together, as noted in the label on Figure 2.11(a).

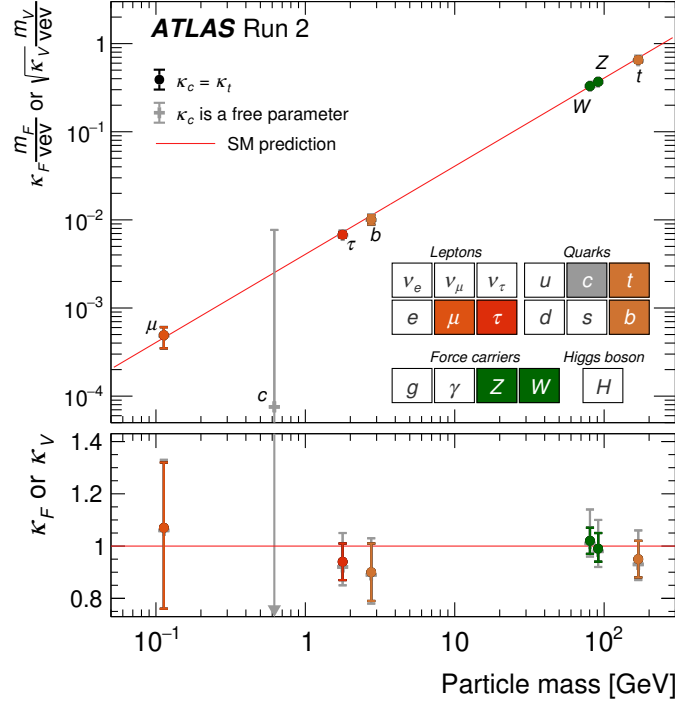


Figure 2.12: Linearity test for the strength of the coupling with the masses for the sundry production and decay channels [32].

Additionally, the SM predicts that the Higgs is a spin 0, CP even⁷. The SM hypothesis has been tested against the other BSM possibilities, and we have evidence that the Higgs is indeed spin 0, and CP even [33].

Finally, the SM predicts that the Higgs width is $\Gamma_H = \frac{\hbar}{\tau_H} = 4.14 \pm 0.02$ MeV, which is much smaller than the $\mathcal{O}(\text{GeV})$ detector resolution. However, the ratio of the on-shell and off-shell cross-sections in $H \rightarrow ZZ$ is proportional to the Higgs lifetime [34, 35]. ATLAS recently used these cross-section measurements to constrain $\Gamma_H = 4.6^{+2.6}_{-2.5}$ MeV at the 68% confidence level [36].

Given this slew of recent results, thus far the Higgs is seeming consistent with the picture built up by the SM. Central to the whole Higgs of the is the ϕ^4 potential elaborated on in Chapter 2.2. So another *crucial* test for the higgs sector is to probe the form of the Higgs potential, $V(\phi)$, for which we need events with two Higgs bosons. Since the Higgs boson has direct interactions with almost all of the SM particles, it's exciting to test this model of electroweak symmetry breaking, and see if the Higgs sector gives clues to some of the big open questions in physics described in Chapter 1. We save a discussion of the phenomenology of the HH signals which probe form of the Higgs potential in Chapter 9.

⁷Symmetric under charge conjugation and parity transformations.

3

The Large Hadron Collider

Every act of creation begins with an act of destruction.

– Pablo Picasso

The Large Hadron Collider (LHC) is a 27 km circumference particle accelerator straddling the border between France and Switzerland, as shown in Figure 3.1 [37]. It is 100 m underground, and accelerates bunches of protons to nearly the speed of light before smashing at the four collision points where the ATLAS, CMS¹, LHCb, and ALICE detectors are located.

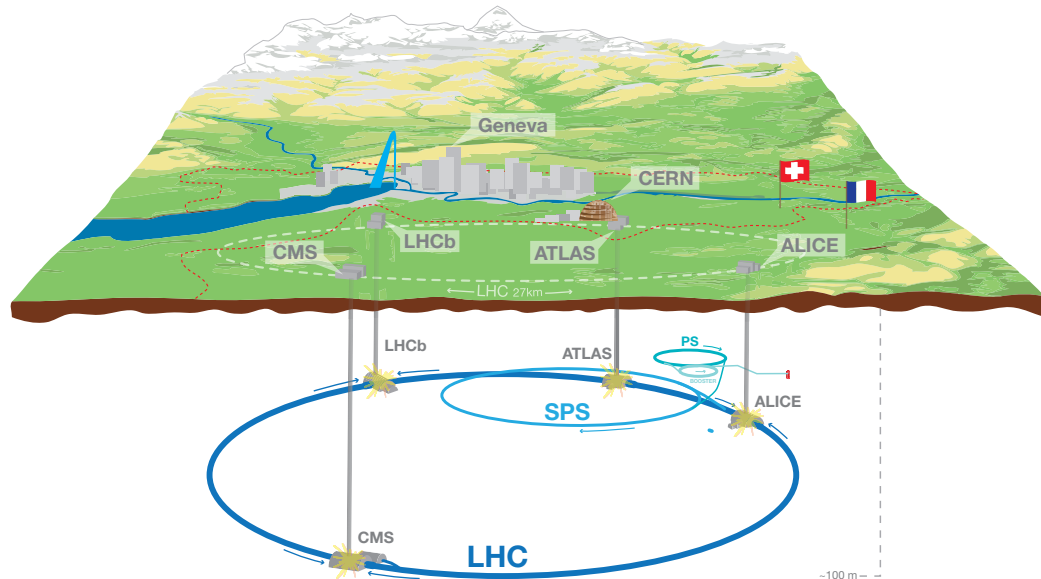


Figure 3.1: Illustration of the LHC [38]

¹ATLAS and CMS are the two general-purpose detectors with similar physics goals: discovering the Higgs boson and searching for BSM particles.

This chapter summarizes the basics of the LHC including the accelerator complex (Chapter 3.1), how the size of the dataset is quantified (Chapter 3.2), and describes relevant data-taking runs. The concept of pile-up is introduced in Chapter 3.3.

3.1 The accelerator complex

The dipole magnetic field in the ring provides the centripetal force to keep the protons moving in a circular trajectory.

$$q\vec{v} \times \vec{B} = -\frac{mv^2}{R}\hat{r} \implies p = qBR \quad (3.1)$$

The acceleration happens in radio-frequency (RF) cavities, in straight segments along the ring. As the beam energy increases, the magnetic field correspondingly increases to keep the protons exactly in the ring. The LHC uses 1232 8 T superconducting dipole magnets to supply the magnetic field for the centripetal force. The dipole magnets are made of NbTi cables and cooled to 1.9 K with superfluid helium to keep them at the superconducting temperature. Higher-order magnets focus the beams for a total of almost 10,000 LHC magnets. A given magnet can only support a $\mathcal{O}(100)$ range in energy, so the sequence of rings is needed to accelerate the protons beam in stages.

The protons come from hydrogen by stripping off the electrons, and then they are accelerated to 50 MeV in the Linac2 [39]. They are then injected into the Booster and accelerated to 1.4 GeV. The Proton Synchrotron (PS) ring increases the energy to 25 GeV, while the Super Proton Synchrotron (SPS) ring accelerates the protons to 450 GeV. Finally, the protons are injected into the LHC ring and accelerated to 6.5 TeV for the dataset in this thesis. The LHC has two bunches of protons with one accelerating clockwise, and the other accelerating counter-clockwise. The circular collider allows the proton beams to use the same ring to accelerate the particles and then recycle the beams to collide the bunches for up to a day before refilling the ring. In the LHC, one proton beam goes clock-wise and the other goes counter-clockwise, and the two beam pipes are immersed in the same magnetic field. Each proton bunch has 10^{11} protons, and the LHC can hold 2808 bunches in each ring.

3.2 Luminosity

The rate is measured by the *instantaneous luminosity*, \mathcal{L}_i :

$$\mathcal{L}_i = \frac{fN_b n_1 n_2}{A}, \quad (3.2)$$

where f is the frequency of a bunch circulating around the ring, N_b is the number of bunches in the ring, n_1 and n_2 are the number of protons in the colliding bunches, and A is an effective area measuring the overlap of the two bunches. The units of the instantaneous luminosity are inverse area

per unit time, and the common unit of area that we use in particle physics is a *barn*: $1 \text{ b} = 10^{-22} \text{ m}^2$.

The size of the dataset is quantified by the *integrated luminosity*, \mathcal{L} , the integral of the instantaneous luminosity over time:

$$\mathcal{L} = \int_0^T \mathcal{L}_i dt. \quad (3.3)$$

Then the number of predicted events is the product of \mathcal{L} and the physics cross-section σ :

$$N = \mathcal{L}\sigma. \quad (3.4)$$

Figure 3.2 sets the scale for the cross-sections relevant at the LHC. At 13 TeV, the total cross-section is $\approx 100 \text{ mb}$ – 18 orders of magnitude larger than the HH signal that the thesis searches for.

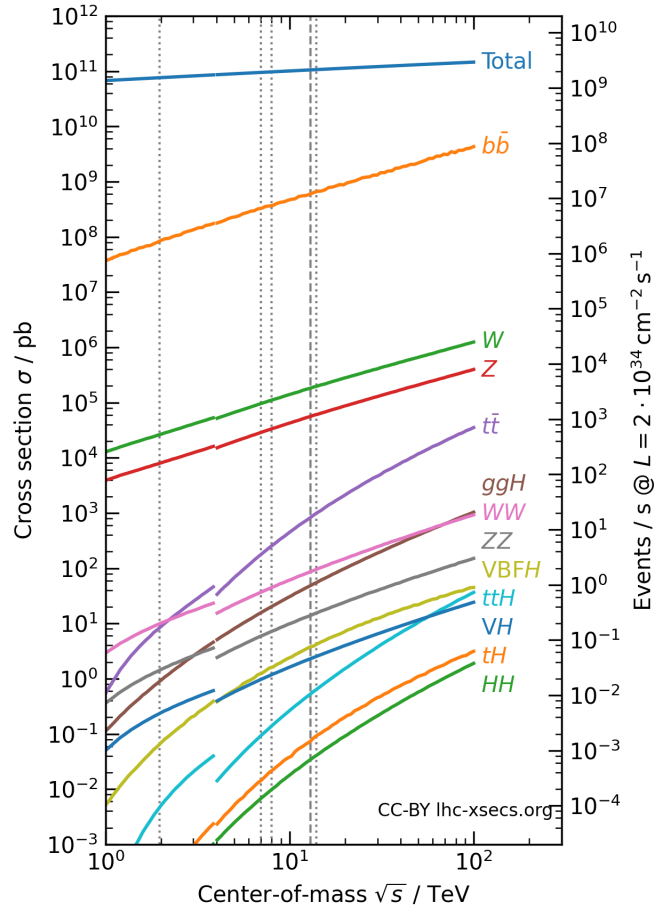


Figure 3.2: Cross sections for some key processes of interest for the Tevatron ($p\bar{p}$) and LHC (pp). The dashed grey line indicates the Run 2 energy, while the dotted lines indicate other past or planned collider energies [40].

Since the physics processes of interest have cross-sections that are so rare compared to the overall event rate, the beam parameters have been modified over the years of data taking to increase the instantaneous luminosity. Figure 3.3 outlines the schedule for the LHC data taking. Data taking for the LHC started in 2011 at an operating energy of 7 TeV, increasing to an energy of 8 TeV in 2012. This was the Run 1 dataset used for the Higgs discovery [10, 11]. The Run 2 period of data taking went from 2015 – 2018 at an energy of 13 TeV, and is the dataset used for this thesis. Although the LHC’s designed peak instantaneous luminosity was $\mathcal{L}_i = 10^{34} \text{ fb}^{-1}$, in Run 2 the instantaneous luminosity was increased to up to twice this design luminosity. The LHC is currently taking data at 13.6 TeV for Run 3, which will double our dataset size. Then in the next seven years after a substantial upgrade to both the detectors and collider, the HL-LHC will collect data at 14 TeV and at a rate of up 7.5 times larger than the design instantaneous luminosity. The HL-LHC will increase the dataset size by another factor of 10.

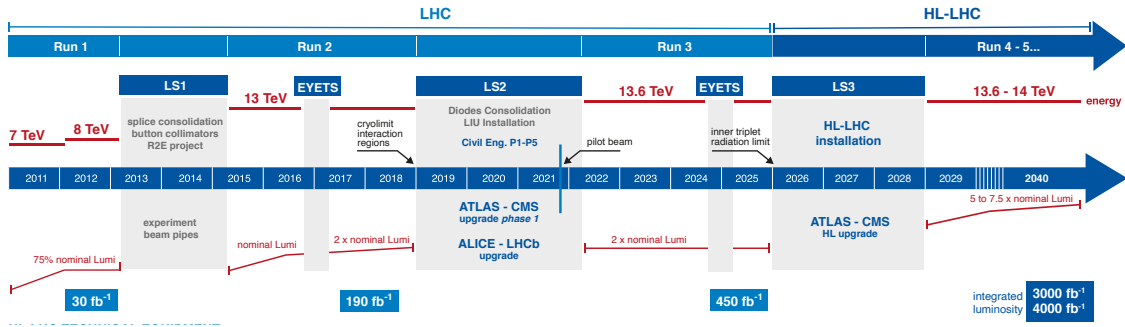


Figure 3.3: Schedule for the LHC data taking [41]. The center of mass energy is indicated by the red horizontal lines above the timeline, while the instantaneous luminosity is shown in the red graphs below.

Figure 3.4 tracks the increase in the integrated luminosity during Run 2. While the LHC delivered $\mathcal{L} = 156 \text{ fb}^{-1}$ of data, it’s only the data that was recorded by the detector with all of the corresponding subsystems operational is usable for physics. Hence, most LHC Run 2 analyses use the main trigger stream with $\mathcal{L} = 139 \text{ fb}^{-1}$.

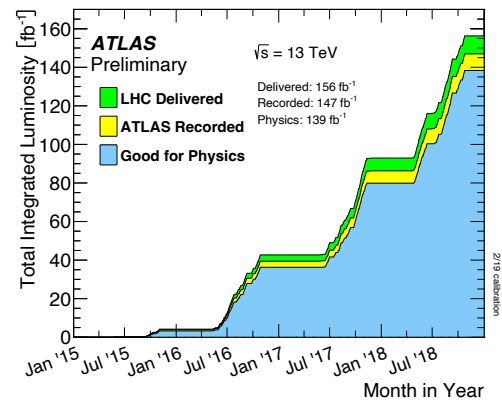


Figure 3.4: Run 2 LHC delivered and ATLAS recorded luminosity [42].

3.3 Pileup

While increasing the instantaneous luminosity does increase the number of signal events, it also increases the number of extraneous collisions, called *pile-up* (PU). There are two types of PU that we deal with. The first is *in-time PU* created from the additional proton-proton collisions in a single bunch crossing. This is characterized by $\langle\mu\rangle$, the average number of collisions per bunch crossing, and Figure 3.5 shows the $\langle\mu\rangle$ distribution in Run 2, and demonstrates how we were able to increase the instantaneous luminosity as we understood the collider and detectors better through the run.

The proton collisions happen every 25 ns, but this is *faster* than the readout time for some of the detector subsystems. The *out-of time PU* refers to the pile-up from other bunch-crossing that can have overlap with the current readout chain.

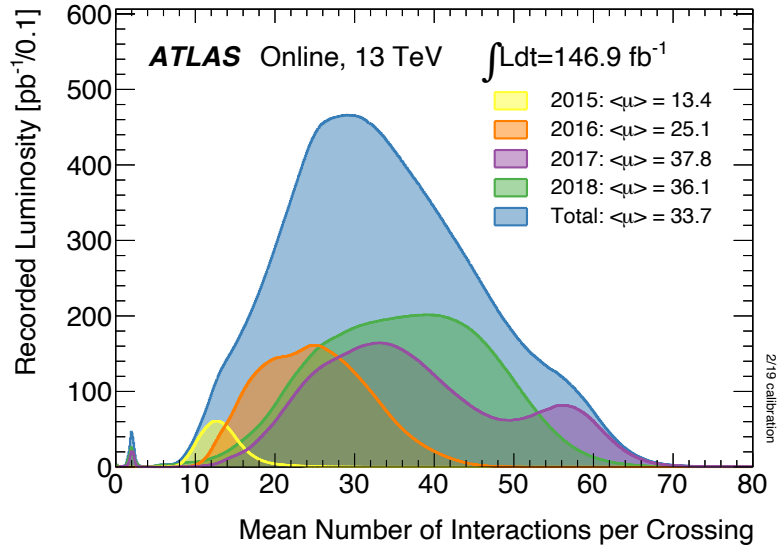


Figure 3.5: PU density profile for Run 2 data taking [42].

4

The ATLAS detector

I never watch the stars – there’s so much down here.

– Lorde “Yellow Flicker Beat”

The ATLAS detector records the data from the proton-proton collisions, and is a 45 m long, 25 m tall cylinder, as shown in Figure 4.1. It has nearly hermetic coverage around the interaction point.

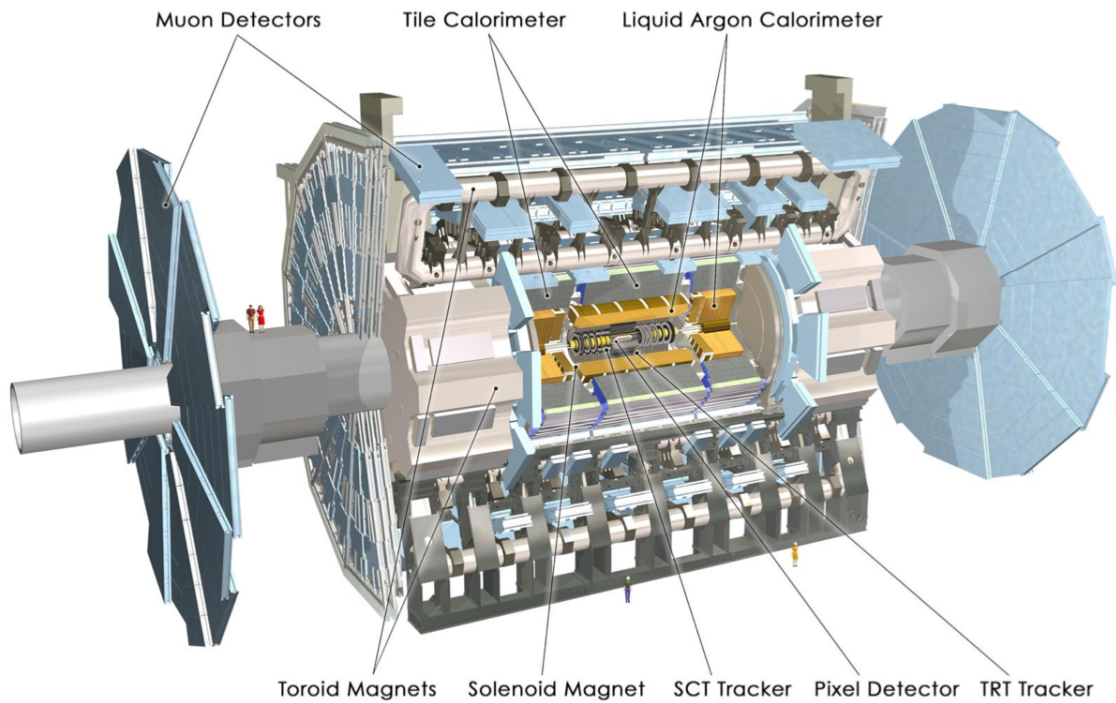


Figure 4.1: The ATLAS detector [43]. The people depicted on the left indicate the detector scale.

The detector elements are organized in layers to give a rich, multi-dimensional view of the reconstructed event. Individual subsystems are optimized to identify the sundry stable particles that come out of the collision, as depicted in Figure 4.2. This chapter enumerates the relevant details of these subsystems for the object identification and subsequent physics analysis most pertinent to this thesis work.

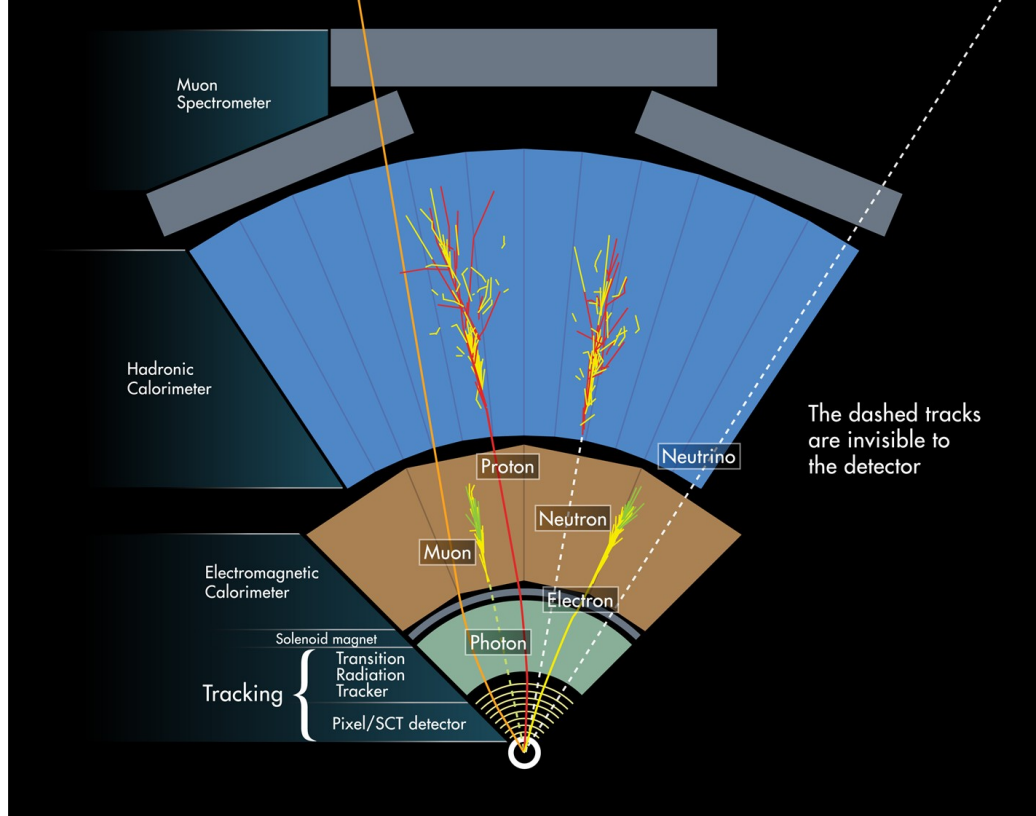


Figure 4.2: Schematic of different detector subsystems optimized for different particles [44].

The inner detector provides charged particle tracking information closest to the interaction point. It is immersed in a 2 T solenoidal magnetic field to bend the trajectories of charged particles and measure the momentum. The subsystems of the inner detector are described in Chapter 4.1 Energy measurement is made in two parts: with an electromagnetic and hadronic calorimeter described in Chapter 4.2. Finally, the outermost layer of the detector is the muon spectrometer, with a 4 T toroidal magnetic field Chapter 4.3. The trigger system decides which events to write to disk, and is summarized in Chapter 4.4.

ATLAS Coordinate System

ATLAS uses a right-handed coordinate system with the z -axis parallel to the proton beam, the x -axis pointing into the center of the LHC, and the y -axis vertically up. Then a cylindrical coordinate system is used. The azimuthal angle, $\phi = \arctan(\frac{y}{x})$, is defined in the plane transverse to the beam, and the pseudorapidity, η , measures the polar angle in the detector:

$$\eta = -\ln \left(\tan \left(\frac{\theta}{2} \right) \right). \quad (4.1)$$

Figure 4.3 shows the relation between θ and η . In the relativistic limit ($m \ll |\mathbf{p}|$ – relevant for the final state stable particles such as electrons, muons and pions), the pseudo-rapidity approaches the rapidity $y = \frac{1}{2} \ln \frac{E+p_z}{E-p_z}$. This is a natural variable for hadron colliders as differences in rapidity are invariant to boosts along the z axis.

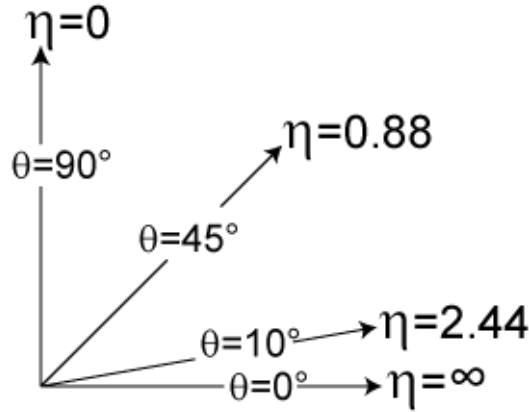


Figure 4.3: Relation between η and θ [45].

As the momentum in the z direction of the colliding partons is not known, momentum is only conserved in the transverse plane. Thus, relevant momentum for object reconstruction is the transverse momentum:

$$p_T = \sqrt{p_x^2 + p_y^2}. \quad (4.2)$$

When the momentum is defined from a calorimeter measurement, E_T (transverse energy) is used instead of p_T . These detector variables are related to four-momentum components using the relations:

$$(E, p_x, p_y, p_z) = (E, p_T \cos \phi, p_T \sin \phi, p_T \sinh \eta) \quad (4.3)$$

$$p = p_T \cosh \eta \quad (4.4)$$

Each of the detector subsystems consists of a cylindrical barrel and the η coverage is extended by including endcaps, which detectors shaped like circular disks oriented transverse to the beamline.

4.1 Inner Detector

Cutaway views of the inner detector (ID) for ATLAS are shown in Figure 4.4. With hundreds of tracks emerging from the interaction point every 25 ns [46], the ID is divided into three different regions to optimize the pattern recognition and momentum measurement algorithms [43]. The pixel detector has the most precision, and is the subsystem closest to the interaction point. The pixel detector also has the highest cost per unit area, so the next least expensive tracking option is the silicon microstrip trackers (SCTs), at the next farthest region away from the interaction point. Finally, the Transition Radiation Trackers (TRTs) compose the outermost region of the inner detector. The inner detector is immersed in a 2 T magnetic field pointing along the z direction and the momentum of the charged particles is measured from the curvature of the tracks in the magnetic field.

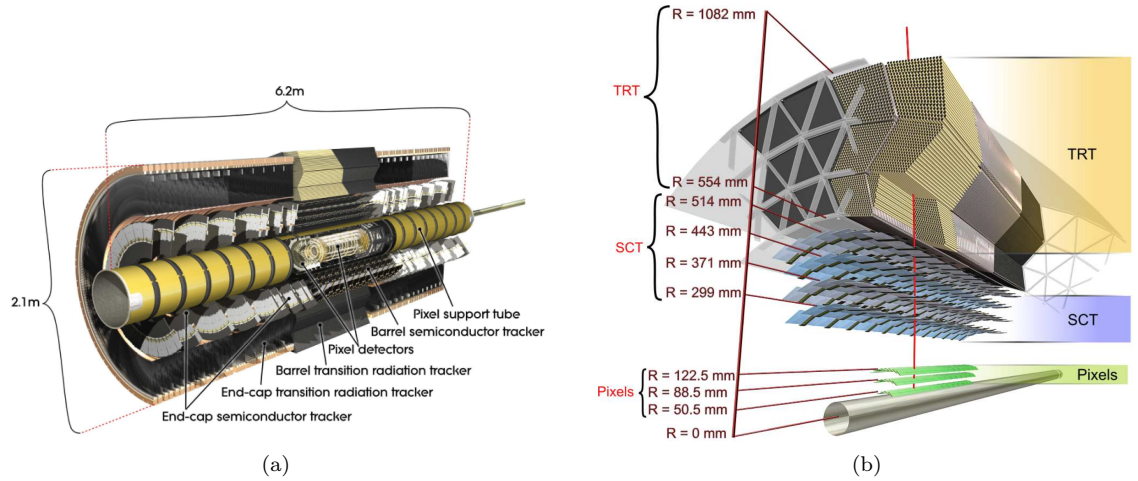


Figure 4.4: Illustration of the subsystems in the ATLAS Inner Detector with the full layout including the endcaps (left) and just the barrel at $\eta = 0.3$ (right) [43, 47]. The IBL is not pictured here as it was installed at the end of Run 1.

Table 4.1 enumerates the positions of the barrel and endcaps layers for each subsystem, and the following subsections go into more detail for each component.

Subsystem	Barrel			Endcap		
	# layers	r [mm]	z [mm]	# disks	r [mm]	z [mm]
IBL	1	33.25	330.15	—	—	—
Pixel	3	50.5 – 122.5	400.5	2×3	88.8 – 149.6	495–650
SCT	4	299 – 514	749	2×9	275 – 560	839 – 2735
TRT	73	563 – 1066	712	160	644 – 1004	842 – 2710

Table 4.1: List of the dimensions of the subsystems in the ID [43, 48].

4.1.1 Pixel Detector

The pixel detector uses the interaction of charged particles in silicon to identify the trajectory, as illustrated in Figure 4.5. As a particle passes through the pixel sensor, it creates electron-hole pairs that are separated by the electric field and accelerated to the metal contacts where the charge is read out by the electronics. The charged particle is not stopped but will continue to leave hits in the subsequent layers of the tracker which can then be used to reconstruct the charged particle trajectory using algorithms described in Chapter 5.1.

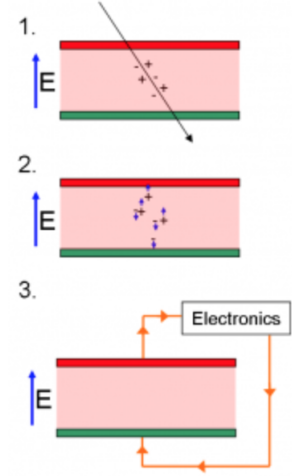


Figure 4.5: How a silicon detector tracks incident particles [49].

The ATLAS pixel detectors use silicon wafers with “ n^+ -in- n ” regions to create the sensors [47], as shown in Figure 4.6(a). The n^+ is an electron-rich region is created by infusing the bulk silicon (which has 4 valence electrons) with nitrogen atoms (which have five valence electrons), and the positive charge comes from the diffusion of electrons away from the region. The “in” region is the intrinsic (or un-doped) region. The other side of the wafer has n -type contacts, which are also created by doping with nitrogen atoms, but with a lower doping level than the n^+ region at the bottom of the wafer. To prevent the flow of electrons away from these n -type contacts, a voltage difference is applied across the two contacts grounding the n -contacts and using a negative voltage at the n^+ contact to create an electric field to keep the electrons in the n -region. The electron-hole pairs are created when an incident particle crosses through the intrinsic region and the signal is read out by the electrons which travel to the grounded n -region contacts.

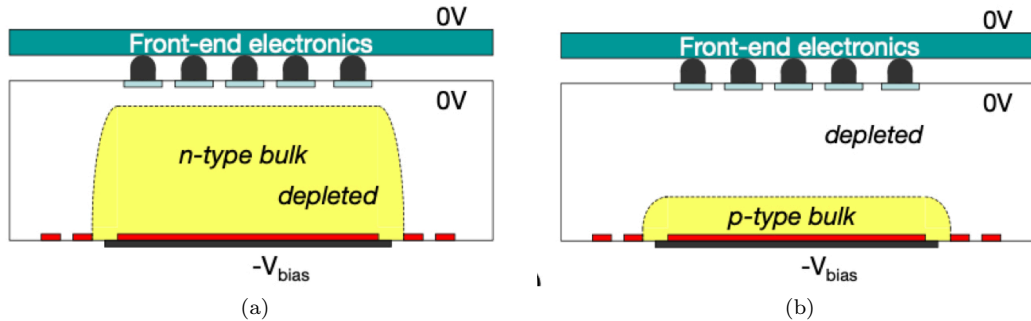


Figure 4.6: ATLAS pixel sensors [47]. Initially they have an n^+ -in- n configuration (left figure), but the irradiation creates a “type-inversion” and subsequently the sensors operate in a p -in- n configuration (right figure). [47]

The holes from the signal are drawn to the n^+ region, and the surplus of holes will eventually balance and become greater than the number of electrons. The n^+ region will turn into a p-type region, a phenomenon called “type inversion”, illustrated in Figure 4.6(b). Since now the sensor has both p and n-type regions, it is now a reverse-biased diode. This type inversion happens after a particle fluency of $\Phi_{eq} \approx 2 \cdot 10^{13} \text{ cm}^{-2}$. As the sensor becomes more irradiated, the size p-type region grows, and once it grows large enough that a conductive channel forms between the two contacts, the sensor can no longer separate electrons and holes to detect incident particles. This rate of charge carrier increase is moderated by using oxygen-infused silicon to create a sensor that will be able to last longer [47].

The original ATLAS pixel detector (from Run 1) has 80 million [50] rectangular $50 \mu\text{m} \times 400 \mu\text{m}$ where the longer dimension of the pixel is in the z direction and the shorter dimension of the pixel is in the transverse direction. These 80 million channels are divided into three cylindrical layers, shown in Figure 4.4(b) and three endcap disks on each side of the cylinder. The pixel detector fits in a 1.442 m long cylinder with a 0.43 m radius [47]. The pixel detector has coverage for $|\eta| < 2.5$ [50]. Since the pixel detector is so close to the interaction point, its electronics must be able to withstand high radiation doses of up to 500 kGy. The p-n diodes can experience a leakage current when an electron-hole pair has enough energy to overcome the potential barrier. The detector is cooled to -6°C to minimize the spontaneous generation of electron-hole pairs.

Between Run 1 and Run 2 of data taking an Insertable B-Layer (IBL) [51, 52] was installed to add an additional layer at a radius of 33.25 mm to cope with the increased luminosity in Run 2. The IBL pixels have a size of $50 \times 200 \mu\text{m}^2$, and adds more than 12 million additional readout channels. The smaller pixel size in the z direction helps improves the longitudinal impact parameter¹ resolution. The proximity to the interaction point mitigates the impact of multiple scattering, which also improves the track’s impact parameter resolution, with the corresponding improvement shown in Figure 4.7. This *directly* improves our ability to tag b -jets, improving our efficiency of rejecting light (charm) jets by up to a factor of four (1.8) [52], with largest improvements at low p_T where the multiple scattering is more impactful. The b -tagging algorithms presented in Chapter 6–7 all benefit from this improved track resolution.

¹The track’s impact parameter is the point of closest approach to a reference point, and will be further elaborated on in Chapter 5.1

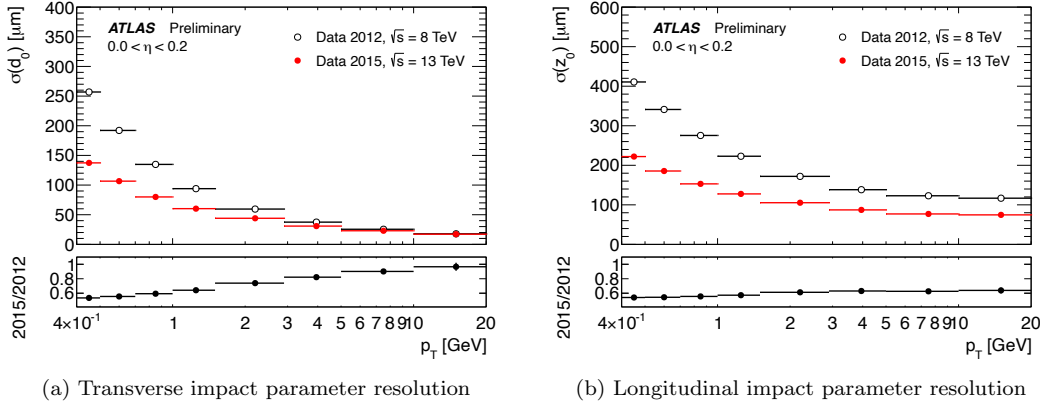


Figure 4.7: Impact of the IBL on the track's momentum resolution, as a function of track p_T . The 2012 data is from Run 1 and does not include the IBL, while the 2015 data from Run 2 shows the lower (better) resolution with the IBL included [53].

4.1.2 Silicon Microstrip Trackers

The Silicon Microstrip Trackers (SCTs) also use silicon technology, except the readout is only in 1d instead of 2d using an $80\mu\text{m} \times 12$ cm strip channel area. The strips are read out on both sides of the silicon modules which are rotated at a stereo angle of 40 mrad relative to each other. These two 2d measurements are combined to recreate the incident particle location. There are four barrel layers and 9 disks in each of the two endcaps with 6.3 million read-out channels. The SCT sensors use p-in-n sensors directly and have an effective operating voltage determined by the effective doping level, and their leakage current also increases linearly with radiation dose. Initially the operating voltage was 150 V, but increased up to 250 to 350 V to compensate for the radiation dose after 10 years of operation.

4.1.3 Transition Radiation Tracker

The Transition Radiation Tracker (TRT) is the outermost region of the ID and composed of 4 mm diameter polyimide drift tubes (straws) filled with gas with an admixture of 70% Xe, 27% CO_2 , and 3% O_2 [43]. In the center of the tube is an anode with a 31 μm diameter cylinder of tungsten wire, as shown in Figure 4.8. The anode is grounded and connected to the front-end electronics, while the inside of the tube are cathodes coated with gold and held at -1.5 kV. Each tube operates as a proportional drift chamber where an incident particle in a straw ionizes the gas molecules to free electrons and create positive ions. These electrons accelerate through the electric field to the anode to create an avalanche as these electrons ionize other atoms with a gain of 2.5×10^4 . Although the silicon trackers need to be cooled between -5 to -10°C , the TRTs operate at room temperature.

The straws are 1.44 m long in the barrel and arranged parallel to the beamline, as shown in

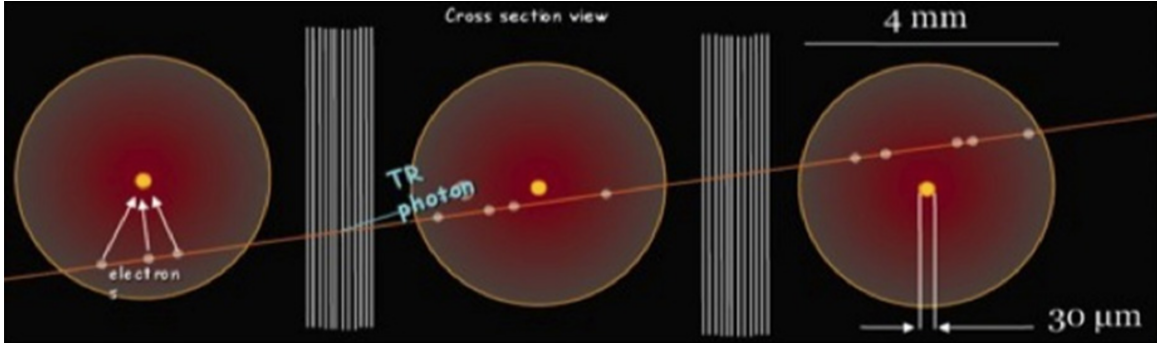


Figure 4.8: Principle of operation for the TRT [54]

Figure 4.4(b). The barrel straws are read out at each end of the tube, and the middle of the tube is insulated with a glass layer to reduce the occupancy of the hits. The straws in the endcap are 37 cm long, and are arranged radially outward, as shown in Figure 4.4(a). The endcap tubes readout the signals at the outer radius of the tube. The TRT has a total of 420 thousand readout channels. The barrel of the TRT only gives (r, z) information while the endcaps only have (z, ϕ) information. An outgoing particle will hit about 36 TRT tubes, and although the TRT hits are less precise than the silicon ones, having more hits improves the momentum resolution.

“Transition radiation” is emitted by a relativistic particle as it traverses a boundary between materials with different permittivities. Electrons emit more transition radiation than pions, and this difference in the transition radiation is measured by the TRT and helps the electron and pion discrimination.

During its lifetime, a straw detector can track 10^{15} particles, and a total integrated charge of 1000 C, which corresponds to 20 years of the LHC’s operation. With the HL-LHC upgrade, the TRT tracker will be replaced with the ITk – an inner detector entirely made of silicon.

4.1.4 Tracker Resolution

The resolution of the tracker is given by:

$$\sigma\left(\frac{1}{p_T}\right) = 0.036\% p_T \oplus 1.3\%. \quad (4.5)$$

Lower momentum tracks have a tighter radius of curvature, so we generally have better momentum resolution for low p_T tracks. However, for tracks with p_T below a GeV, the multiple scattering will start to degrade the track resolution.

4.2 Calorimeter

Calorimetry literally means “heat measurement.” The incident particle interacts with the material in the detector to form a shower of particles that are decelerated and absorbed in the detector material to reconstruct the original particle’s energy. The ATLAS calorimeter (shown in Figure 4.9) is divided into two parts: the electromagnetic calorimeter (ECAL) and the hadronic calorimeter (HCAL). The ECAL has higher precision than the HCAL and is designed to measure (and contain) the electromagnetic interactions. The hadronic interactions happen in both ECAL and HCAL, but longer interaction length for the particles interacting by the strong force means their showers are generally stopped in the HCAL.

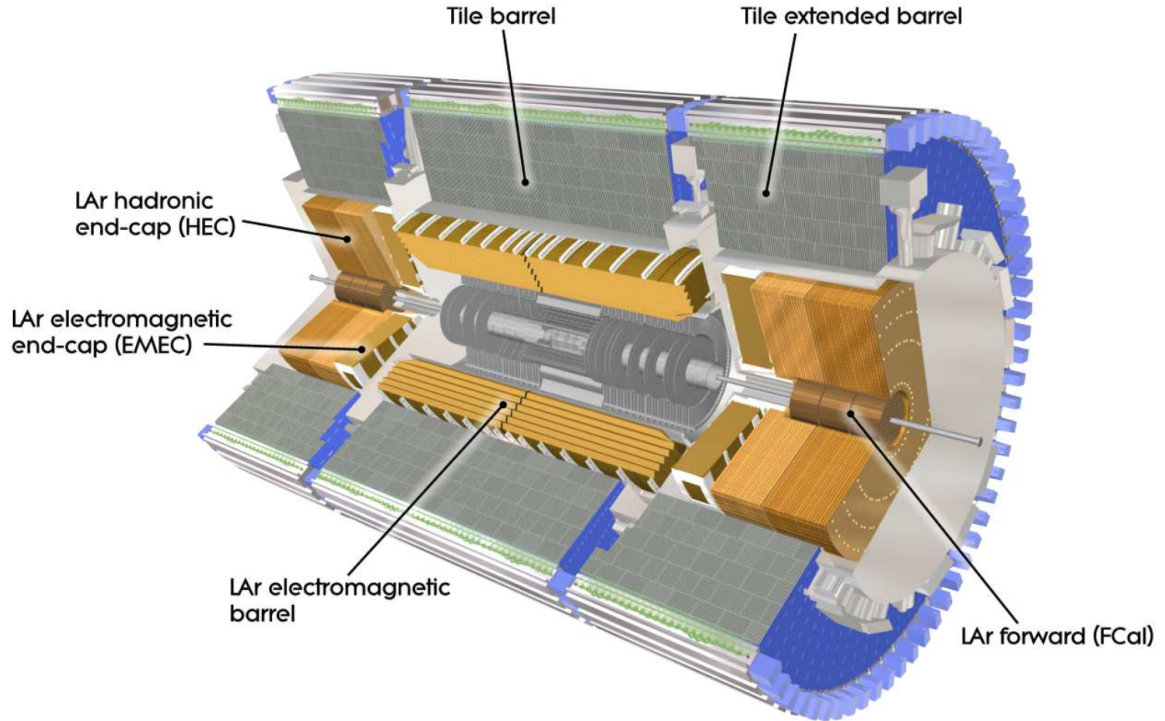


Figure 4.9: Calorimeter system at ATLAS [43].

4.2.1 ECAL

EM shower

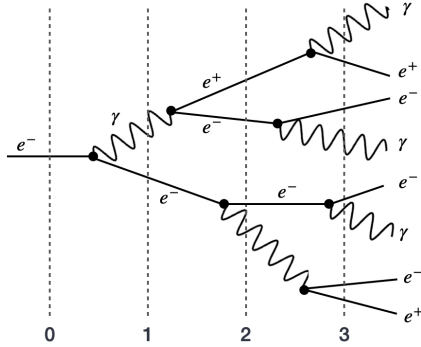


Figure 4.10: The development of an electromagnetic shower [55]. The dashed lines and numbers indicate the number of radiation lengths, X_0 .

The radiation length is given by:

$$X_0 = \frac{716.4 \text{ g} \cdot \text{cm}^{-2} A}{Z(Z+1) \ln \frac{287}{\sqrt{Z}}}, \quad (4.6)$$

where A is the number of nucleons and Z is the atomic number of the detector material [56].²

A sampling calorimeter divides the calorimeter into absorber and active slabs, shown in Figure 4.11. The absorber slabs use a high Z material with a smaller radiation length so the shower develops more quickly. The energy deposited in the active region of the calorimeter is measured, and the total energy is the energy deposited in the absorbing region divided by a scale factor, $f_{\text{sampling}} = E_{\text{visible}}/E_{\text{deposited}}$. A sampling calorimeter contains the shower in a smaller detector to help minimize the cost of the experiment. However, the sampling method is not as precise as only a fraction of the energy is measured, and fluctuations proportional to \sqrt{E} for Poisson statistics introduce extra errors into the measurements.

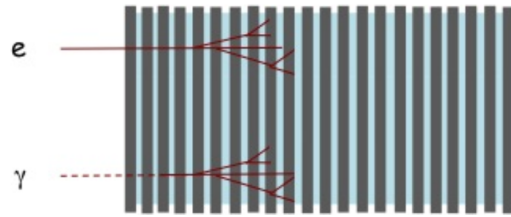


Figure 4.11: Schematic illustrating how a sampling calorimeter causes an incident particle to form a shower more quickly. [58]

²The units of X_0 are $[\text{density}] \times [\text{length}]$, so dividing by the density corresponds to the particle's distance traveled in the material before undergoing an interaction [57].

Liquid Argon Detector

The ATLAS ECAL is a sampling calorimeter arranged in an accordion structure. Figure 4.12 justifies this accordion shape design choice. If the sampling slabs were perpendicular to the incident particle – which allows us to accurately measure the particle energy. However, the ECAL cylinders are very large (6.4 m long, with an outer (inner) radius of 4 m (2.8 m)). The long cables needed to read out the signals from the sides of the detectors add an extra 10 – 20 ps to the signal collection time, which would degrade the resolution due to out-of-time pile-up (PU). Alternatively, placing the plane parallel to the incident particles would give very poor energy resolution if the particle happened to hit the absorber material of the sampling layer. Thus the accordion structure presents a goldilocks solution that minimizes the cable length while preserving the reconstructed energy resolution.

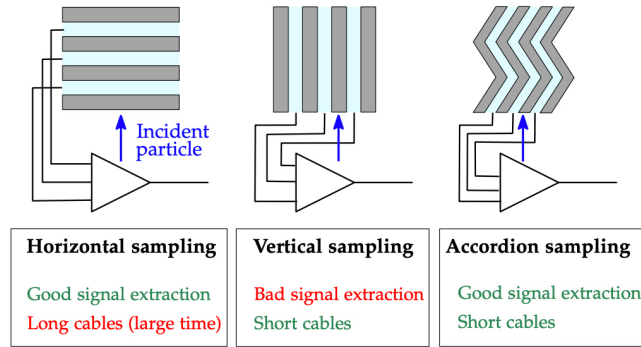


Figure 4.12: Motivation for the accordion structure [59]

The high Z material ($Z = 82$) lead creates the shower, while the energy is measured in liquid argon (LAr), a low Z material ($Z = 18$). The LAr scintillators are read out with wavelength-shifting photosensors. The calorimeter is arranged in a tower structure with modules forming wedges pointing back to the interaction point, shown in Figure 4.13. An electron or photon shower is contained within a few modules or cells.

There are two calorimeter sections: the barrel covering $|\eta| < 1.475$ and the endcaps covering $1.375 < |\eta| < 3.2$. Each of these sections has four layers: a pre-sampler (which helps discriminate between electrons and pions), and three sampling layers. The segmentation in the barrel is $\Delta\eta \times \Delta\phi$ is 0.025×0.1 in the pre-sampler region, 0.0031×0.1 in the strips, 0.25×0.25 in the main region, and 0.5×0.025 in the back region. This gives an energy resolution of the ECAL of:

$$\frac{\sigma(E)}{E} = \frac{10\%}{\sqrt{E}} \oplus 0.7\% \oplus \frac{0.1\%}{E}. \quad (4.7)$$

The first term ($\propto 1/\sqrt{E}$) is the stochastic noise from the randomness in the number of particles created in the shower and absorbed in the active material. The second (constant) term is due to

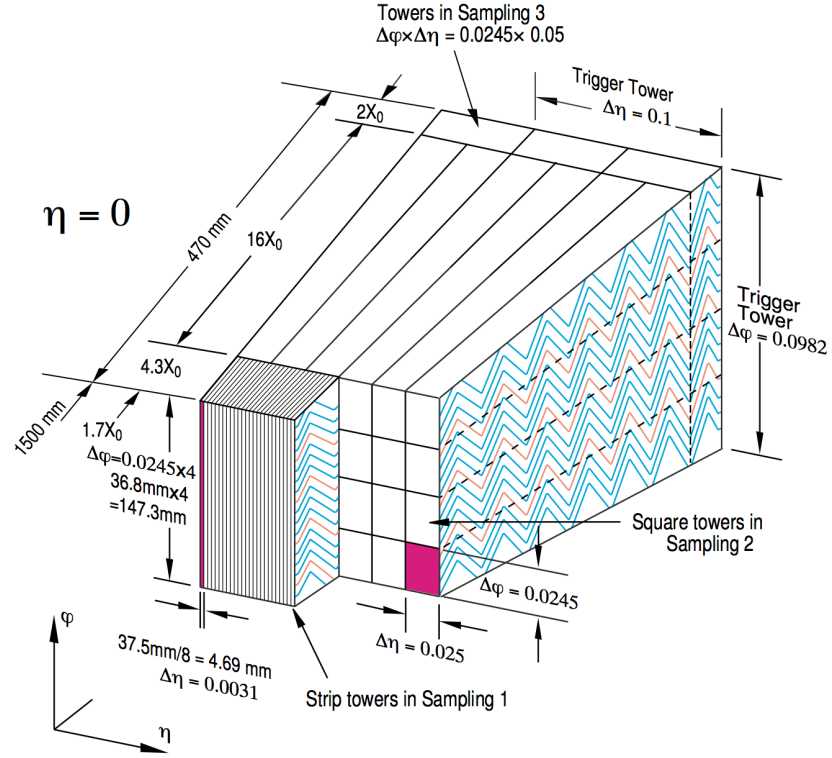


Figure 4.13: Wedge with the accordion structure in the LAr calorimeter [60].

calibration errors and the inactive material present. The last term ($\propto 1/\sqrt{E}$) is the noise term due to the underlying PU and the electronics noise.

PU is more relevant in the calorimeter as the signal collection time is 450 ns, which is a factor of 18 longer than the bunch crossing rate of 25 ns. To mitigate the impact of pile-up, the calorimeter signals are reshaped, as shown in Figure 4.14. The integral of a signal pulse is zero to give a first-order cancelation of the impact from out-of-time pile-up (described in Chapter 3.3). The bunch crossing of the signal can be deduced by fitting the sampled pulse shape.

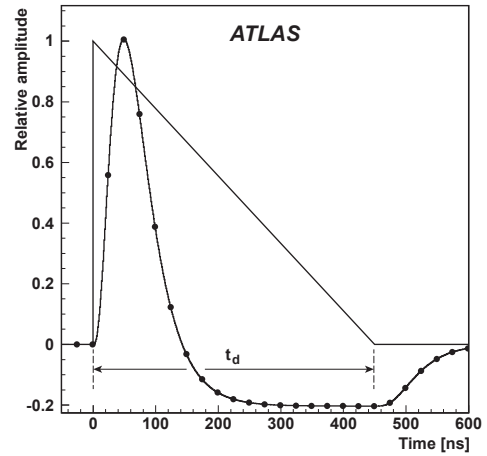


Figure 4.14: Signal in the EM barrel calorimeter [61].

4.2.2 HCAL

Hadronic showers develop similarly to the EM, but are characterized by the interaction length, $\lambda = 37.8A^{0.312} \text{ g} \cdot \text{cm}^{-2}$, the average distance that a particle travels before undergoing a nuclear interaction. Since λ is generally 5-10 times larger than X_0 the hadronic showers take longer to develop than the EM ones. The hadronic calorimeter (HCAL) is the layer just outside of the ECAL, and measures the energy of the hadron shower. It is divided into three parts: the tile calorimeter, the LAr hadronic end-cap calorimeter, and the LAr forward calorimeter, also shown in Figure 4.9.

As hadronic showers have higher intrinsic response fluctuations, the resolution of HCAL is coarser. While the ECAL has over 180 readout channels, the HCAL adds only about 14 thousand readout channels. The HCAL has an energy resolution of [62]:

$$\frac{\sigma(E)}{E} = \frac{50\%}{\sqrt{E}} \oplus 3.4\% \oplus \frac{1\%}{E}, \quad (4.8)$$

a factor of 5–10 worse than the ECAL energy resolution (from Eq. 4.7). The reconstructed energy from the hadronic calorimeters is important for reconstructing quark and gluon p_T , as will be described in Chapter 5.3. The components of the HCAL are summarized in the following.

Tile Calorimeter

The tile calorimeter is just outside ECAL and has an inner radius of 2.28 m and an outer radius 4.45 m. The barrel covers $|\eta| < 1.0$, while the extended barrels cover $0.8 < |\eta| < 1.7$. It is a sampling calorimeter with a steel absorber and scintillating tiles read out with wavelength-shifting fibers and photomultiplier tubes, as shown in Figure 4.15. Azimuthally divided into 64 modules, the barrel is segmented into three regions at 1.5λ , 1.8λ , and 4.25λ , while the extended barrel has 1.5λ , 2.6λ , and 3.3λ depth segmentation. At $\eta = 0$, the HCAL has a depth of 9.7λ [43].

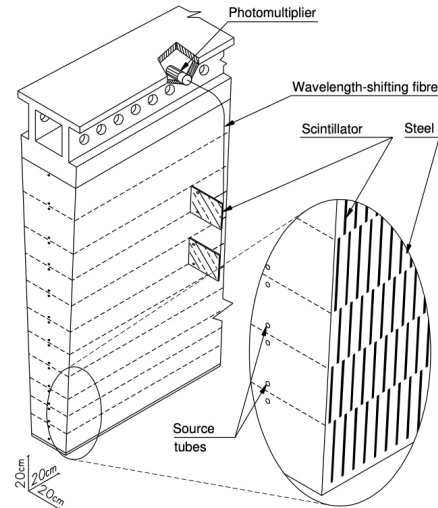


Figure 4.15: An element of the tile calorimeter [43].

LAr hadronic end-cap calorimeter

The LAr hadronic endcap calorimeter has two wheels per endcap just behind the ECAL endcap and is housed within the same LAr cryostats. The wheels have two depth segments, 25 mm parallel copper plates for the wheels closest to the interaction point, and 50 mm plates for the wheels farther from the interaction point. The wheels are divided into 32 wedges, with inner and outer radii of

0.475m and 2.03m, respectively. The copper plates are used as the absorbing material, and are filled with LAr as the active sampling volume [43].

LAr forward calorimeter

The forward calorimeter covers $3.1 < |\eta| < 4.9$ and is approximately 10 interaction lengths deep. Each side has three modules, the first of copper for extending the ECAL, and the second two of tungsten for the HCAL.

4.3 Muon Spectrometer

Although muons interact and deposit some of their energy in the calorimeter, they are minimum ionizing particles and are not absorbed by the calorimeter and the muon spectrometer adds additional hits for the muons to improve the muon track precision. The muon momentum is measured by the deflection of their tracks in a 4 T toroidal magnetic field [43]. The muon detector is a gas detector like the TRT, and operates according to similar principles.

Monitored Drift Tubes

The Monitored Drift Tubes (MDTs) are drift chambers that provide precision measurements of the muon track. Each tube is made of aluminum with a 3 cm diameter (shown in Figure 4.16(a)) and length between 0.9 and 6.2 m. The tube is filled with a 93% Ar, 7% CO₂ gas mixture, at a pressure of 3-bar [43]. It has a gain of 2×10^4 [43], similar to the TRT drift chambers (see Chapter 4.1.3). To control the precision, the sag of the wires is minimized by three kinematical mounts to minimize distortion due to the support, as shown in Figure 4.16(b). There are 1174 MDTs in the barrel of ATLAS ($|\eta| < 2$).

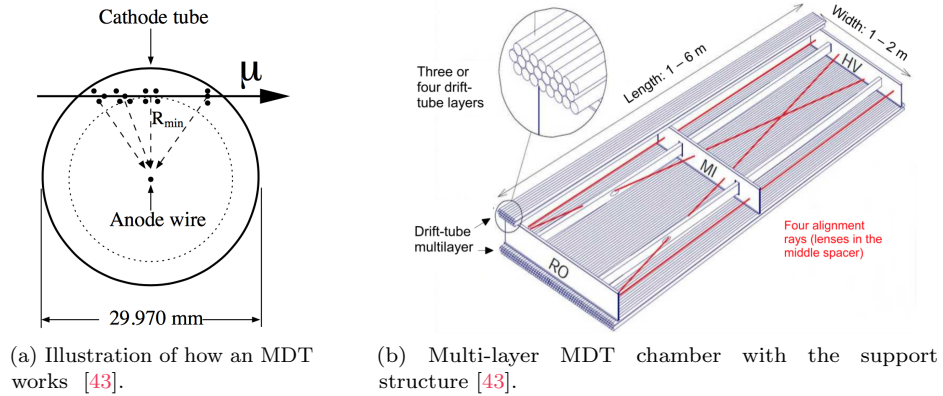


Figure 4.16: Monitored Drift Tubes: MDTs

Cathode Strip Chambers

For larger pseudorapidities ($2 < |\eta| < 2.7$), Cathode Strip Chambers (CSCs) are used for their higher granularity in the endcap region [43]. This is a multi-wire proportional chamber with strip readout with a sense wire pitch of 2.54 mm and a read-out strip pitch of 5.08 mm resulting in a $60 \mu\text{m}$ track resolution [43].

The MDTs and CSCs together provide almost 40 thousand extra readout channels to give high-precision hits for offline track reconstruction.

Muon Trigger Subsystems

The hardware-based muon trigger identifies muons with high p_T pointing back to the interaction point [43]. Different subsystems are used for different regions of the muon detector to account for the varying muon rates. In the barrel region, **Resistive Plate Chambers** (RPCs) are used. They're operated in avalanche mode, which allows large current outputs to be generated by a slight current input. Each chamber has two gas-volumes read out with four planes of read-out strips to provide a read-out on each end for each gas volume. The detector uses two RPCs surrounding the innermost MDT layer tuned to detect a low p_T threshold, and then a third outermost RDC tuned to detect a high p_T threshold as illustrated in Figure 4.17. The time-resolution of these devices is 2 ns. In the endcaps, **Thin Gap Chambers** (TGCs) are operated in the saturation mode, which gives a constant gain per unit of input current. The low and high momentum resolutions are $\approx 30\%$ and $\approx 20\%$, respectively. The low p_T muon measurement is limited by the multiple scattering and energy fluctuations, while the high p_T muon resolution is limited by the interaction region's length.

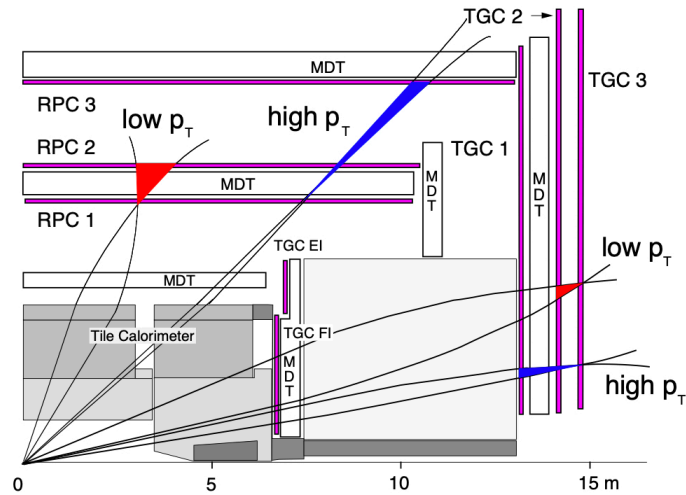


Figure 4.17: Muon trigger system [43].

4.4 Trigger system

Collisions every 25 ns give a data rate of more than 40 TB per second, and it is practically infeasible to save all of these events. ATLAS utilizes a two stage of trigger that determine which events to save and write to disk [43], with an overview in Figure 4.18.

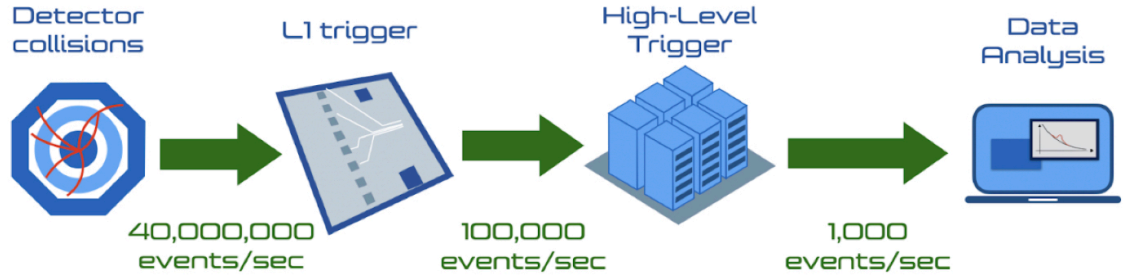


Figure 4.18: Illustration of the data acquisition chain – a sequence of decisions in hardware (L1) and software (HLT) which determines which events get saved [63].

The first stage is the Level I (L1) trigger takes a very coarse view of the detector and has a μs to decide whether to keep the event. This prohibits computationally expensive algorithms (such as tracking) being run, so L1 only looks at trigger objects using information from the calorimeter and muon systems. Since these algorithms need to be very fast, they’re implemented in hardware using field programmable gate arrays (FPGAs) before the electronic signals are read off of the detector. It reduces the rate by almost three orders of magnitude (to 100 kHz) for sending the electronic signals off of the detector.

The second stage is High Level Trigger (HLT) is the software based trigger system, that implements an event reconstruction very similar to the offline algorithms. At this stage, tracking information is available, so b -tagging information can be used in the trigger decision (of particular importance for the $HH \rightarrow 4b$ analysis). There are several different “trigger streams” or combinations of selections applied to the trigger objects to collect datasets of interest to the diverse physics program. Events passed from all of these trigger steams get saved at a rate of 1 kHz while data is being taken.

5

Event Reconstruction

The things which are seen are not made of the things which do appear.

– Hebrews 11:3

The task of event reconstruction involves taking electronic read-out signals from the 100 million sensors in the detector to reconstruct the physics objects to use in high-level physics analyses. Since the LHC collides protons, we produce lots of quarks and gluons in the final state, with an example of the collimated spray of particles, or jet, produced in the detector shown in Figure 5.1. The $HH \rightarrow 4b$ analysis in this thesis also has a final state with a large number of jets, so this chapter emphasizes the reconstruction of jets and the tracks that are most important for this thesis work.

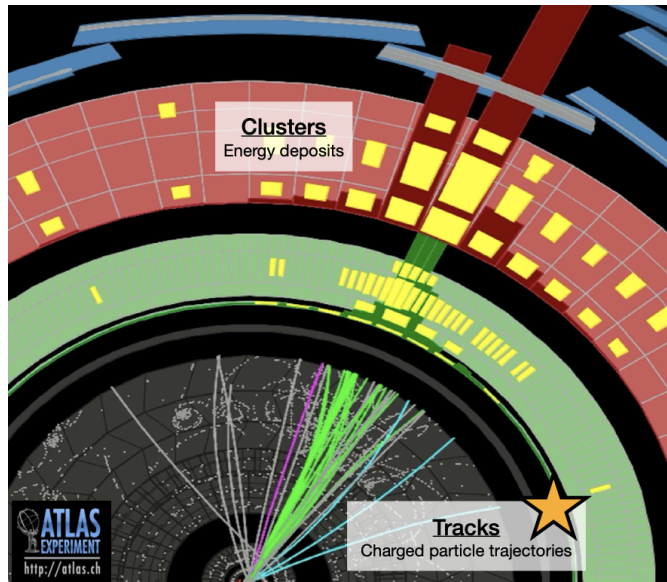


Figure 5.1: Illustration of what a jet looks like in our detector [64].

This chapter is organized as follows. Chapter 5.1 discusses the reconstruction of the charged particle trajectories in the inner detector. The tracks that originate from the same location are then reconstructed into vertices, as described in Chapter 5.2. The unsupervised learning algorithms that reconstructs the quark and gluon decay products into the “jets” are described in Chapter 5.3. Finally, we conclude with the reconstruction of other objects less directly related to the $HH \rightarrow 4b$ analysis: electrons and photons in Chapter 5.5.1, muon tracks in Chapter 5.5.2, hadronic τ decays in Chapter 5.5.3, and the missing energy in Chapter 5.6.

5.1 Tracks

Inner detector (ID) tracks are reconstructed from the hits (or charge depositions) in the pixel, SCT, and TRT detectors. With a set of spatial track hits, the parameters of the track can be defined with a Kalman filter [65, 66]. A Kalman Filter extends the Hidden Markov Model for a continuous vector space for the hidden state. The term *Markovian* means the full state of the system is encapsulated by the track’s state, which is given in the perigee representation:

$$\left(\frac{q}{p}, \theta, \phi, d_0, z_0 \right), \quad (5.1)$$

where the first three variables are the track’s momentum in polar coordinates, and the d_0 and z_0 are the **impact parameters** which define how this track is displaced from a reference point, which is either the center of the beam spot or the “primary vertex” (defined in Chapter 5.2). These track parameters in three dimensions are illustrated in the graphic, Figure 5.2

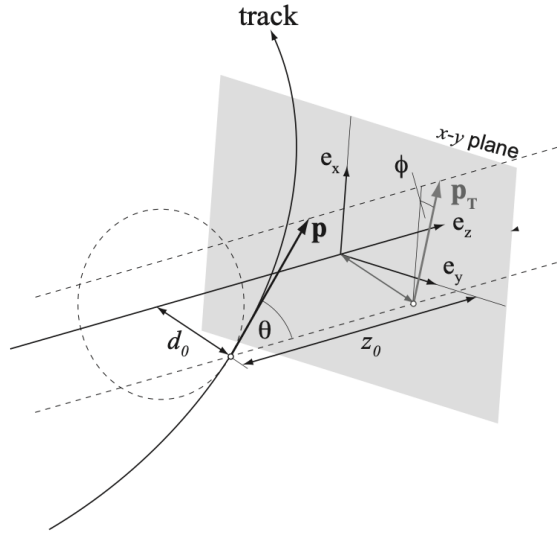


Figure 5.2: Visualization of the track helix with the perigee parameters indicated [67].

The Kalman Filter fits these perigee parameters minimizing the residuals, or the mean squared error between the measurements and the prediction. It is a recursive estimator, so we iteratively add measurements to the state vector, and also remove outliers. The fit returns both the perigee parameters and the associated errors.

We also often consider two transverse projections which are shown in Figure ???. The relevant variables on the graphic are also enumerated below.

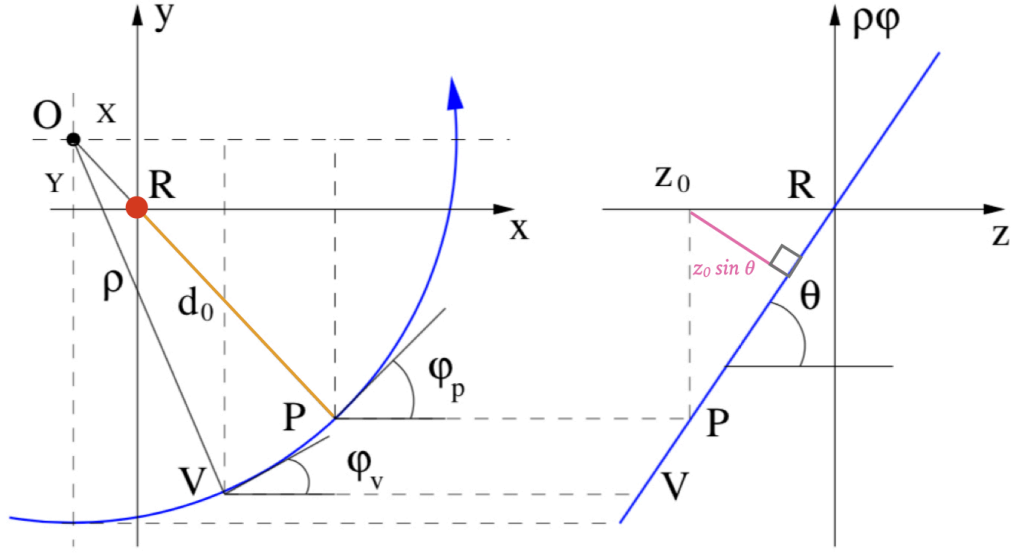


Figure 5.3: The track in the transverse (left) and $\rho\phi$ (right) planes. The track in the perigee representation is defined at the point of closest approach, P, relative to a reference position, R [68].

- **R**: reference position at which the tracks are defined, i.e, for b -tagging we use the primary vertex of the collision as the reference point.
- d_0 : The transverse impact parameter, point of closest approach (POCA) in the transverse plane with respect to R.
- $z_0 \sin \theta$: The longitudinal impact parameter, or the longitudinal displacement from the POCA (defined in the transverse plane). The multiplication by $\sin \theta$ term is included because it characterizes the 2d distance from z_0 to the closest point along the track trajectory.

5.1.1 Track reconstruction

The “hits” of the track are the space points used to reconstruct the track’s trajectory, as indicated by the yellow dots in Figure 5.4. The pixel clusters directly give three-dimensional space points. The strips readout is a 2d position, so the two strips from the front and back of the module need to be combined with the 40 mrad stereo angle to give a 3d space-point. The TRT readout gives a 2d space point with (r, ϕ) in the barrel and (r, z) in the endcaps.

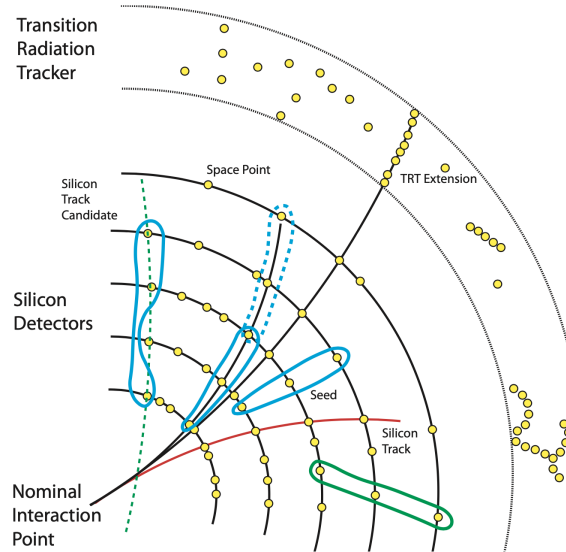


Figure 5.4: Graphic for the ATLAS inside-out track reconstruction steps [69].

The primary ATLAS track reconstruction algorithm is the **inside-out** algorithm with starts searching for hits in the silicon detector [70], with the steps enumerated below:

- *Seed finding*: Triplets of hits in the silicon detector are used to seed the track reconstruction. The track seeds are indicated by the blue circles in Figure 5.4. With these three hits and assuming a uniform magnetic field, a first estimate of the track parameters is obtained.
- *Hit collection*: The track is extrapolated into the other silicon layers and the hits are collected. This is shown by the dashed blue circle in Figure 5.4.
- *First track fit*: A combinatorial Kalman filter uses these hits to refine the estimate of the track parameters which is the Silicon Track Candidate. At this stage, there can be several track candidates for each track seed.
- *Ambiguity solving*: A collection of NNs which decides whether a given cluster is shared between two tracks and how to split the energy deposition between these multiple tracks [71]. We will elaborate more on this step in the following section, Chapter 5.1.2. The result of this step is the “silicon track”, shown in red in Figure 5.4.
- *TRT extension*: The silicon track is extended to the TRT and the compatible TRT hits are included to improve the track’s momentum resolution.

To improve the efficiency for tracks from decays displaced from the original collision point, an **outside-in** track reconstruction algorithm is also used [70]. The track is seeded with hits from the TRT. The track is extended to include hits from the silicon detector, with an ambiguity solver again applied to mitigate the hit sharing between two tracks.

5.1.2 Challenges in Dense Environments

A single charged particle can deposit its energy in multiple pixels due to a few different effects [71]: (1) an incidence angle for the particle, (2) the diffusion of the electrons and holes as they drift to the read-out electrodes, and (3) the drift of the electrons caused by the B-field. A track cluster is between 1.4 – 3 pixels in the transverse direction and 1 and 3.5 pixels in the longitudinal direction [71].

In the dense environment at the core of the jet, the corresponding charged particles will be more collimated, and it can become more difficult to reconstruct the trajectories as the charge depositions are overlapping. Figure 5.5(a) shows how the energy depositions of the charged particles can make it difficult for the track reconstruction algorithm to reconstruct the individual particle trajectories. Additionally, more energetic jets will have more fragmentation tracks, thus decreasing the distance between the tracks becomes smaller. Figure 5.5(b) shows the average distance between the closest two particles in the jet as a function of jet p_T . For jets with $p_T > 200$ GeV, the average longitudinal separation is *smaller* than the size of the pixel detector¹. The pixel feature size is smaller in the transverse direction, but jets with $p_T > 1$ TeV will also have constituents closer than the pixel size.

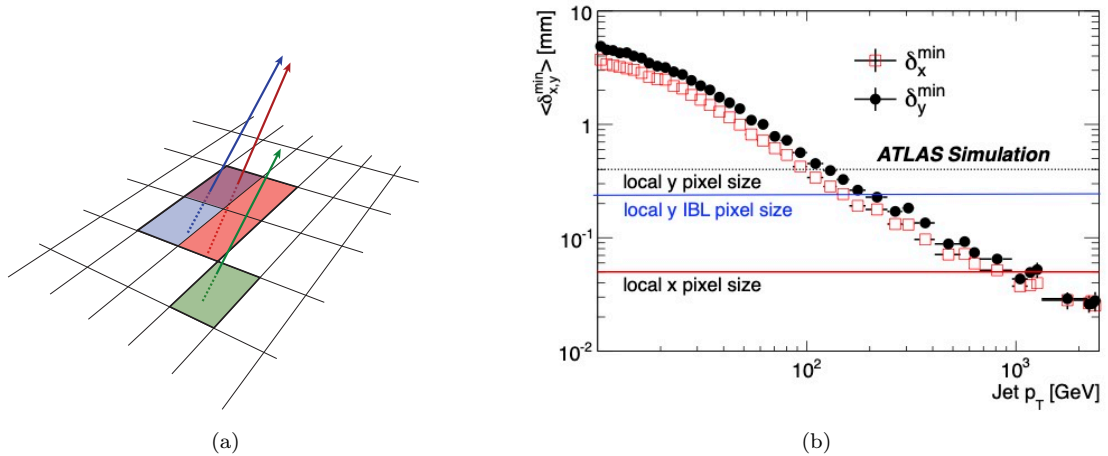


Figure 5.5: Illustration of the challenges for reconstructing the charged particle trajectories in the dense environments at the core of jets [71]. **Left:** In the core of a jet, several charged particles can be reconstructed in a single cluster. **Right:** The average separation (both in the transverse, x , and longitudinal, y directions) of the closest stable particles in jets as a function of jet p_T . Also shown on the figure is the pixel size for the standard pixels and the IBL pixels.

To enhance the track reconstruction for high p_T jets, ATLAS uses a set of NNs to (1) determine if the cluster is more likely to be due to more than one incident particle and then (2) estimate the positions and errors of the clusters. A cluster (or hit) can be “shared” between several tracks, and if

¹The IBL pixels are $200 \times 50 \mu\text{m}^2$, and the subsequent pixel layers are $400 \times 50 \mu\text{m}^2$ (see Chapter 4.1.1)

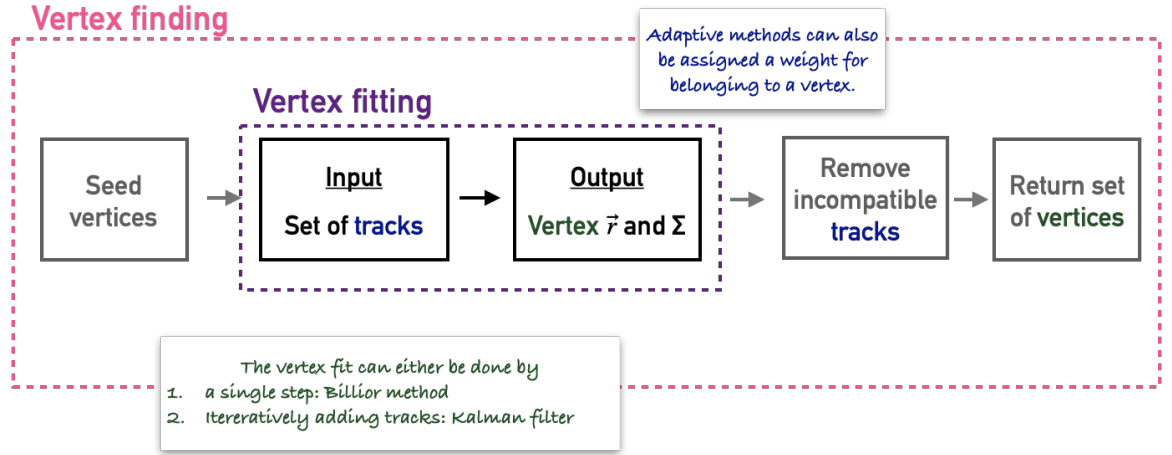
the probability is high enough the charge deposition of the cluster is split to give multiple “hits” to provide separate track measurements for the final track fit.

5.2 Vertexing

5.2.1 General problem formulation

This formulation of the vertexing problem follows from the discussion in Section 5.3 of [68].

The task of vertex fitting is the pattern recognition task of *identifying* which tracks came from a common origin point. This task has (broadly speaking) two steps, which are illustrated below.



The step of **vertex finding** involves the pattern recognition task of determining which tracks might come from the same origin. The **vertex fit** then estimates the position and corresponding covariance matrix for the intersection of the tracks in a common vertex. The contribution of a track to the vertex error can be used to assess the compatibility of each track with the fitted vertex. Thus, these two steps are tightly connected. We’ll begin by detailing the a vertex fit.

From the perigee fit, each track has a measure of its error, and the location of the track in space can be specified by the *Gaussian probability tube*:

$$P(\vec{r}) = \int d\phi_p \exp \left[-\frac{1}{2} (\vec{r} - \vec{r}(\phi_p))^T \text{COV}_{3 \times 3}^{-1}(\phi_p) (\vec{r} - \vec{r}(\phi_p)) \right], \quad (5.2)$$

where $\text{COV}_{3 \times 3}^{-1}$ is the inverse projection of the 5x5 covariance matrix in the track perigee representation into the 3x3 subspace represented by the Cartesian coordinates, ϕ_p characterizes the location along the extrapolated track.

Given these individual track probability tubes, we can assess the combined probability by

integrating over the probabilities of all N of the tracks:

$$P(\vec{r}) = \int d\phi_{p,1} d\phi_{p,2} \dots d\phi_{p,N} \prod_{i=1}^N \exp \left[-\frac{1}{2} (\vec{r} - \vec{r}_i(\phi_{p,i}))^T COV_{3 \times 3,i}^{-1}(\phi_{p,i}) (\vec{r} - \vec{r}_i(\phi_{p,i})) \right], \quad (5.3)$$

where \vec{r} is vertex location maximizes the likelihood of Eq. 5.3, and $\vec{r}_i(\phi_{p,i})$ is track i at the point $\phi_{p,i}$ along the trajectory [68].

The integral above is only non-zero when all of the terms are non-zero. Each track is a sparse representation (zero for most of the space), thus the integrand does not have a strong dependence on the product, and the integrands factorize. Then we can evaluate of the integrals. As each integral is only non-zero for a very small range of $\phi_{p,i}$, we replace it with $\hat{\phi}_{p,i}$ the point of closest approach of the track to the proposed vertex location \vec{r} :

$$P(\vec{r}) = \prod_{i=1}^N \exp \left[-\frac{1}{2} (\vec{r} - \vec{r}_i(\hat{\phi}_{p,i}))^T COV_{3 \times 3,i}^{-1}(\hat{\phi}_{p,i}) (\vec{r} - \vec{r}_i(\hat{\phi}_{p,i})) \right]. \quad (5.4)$$

Finding the \vec{r} that maximizes $P(\vec{r})$ is equivalent to finding the \vec{r} that minimizes the χ^2 test statistic:

$$\chi^2(\vec{r}) = \sum_{i=1}^N (\vec{r} - \vec{r}_i(\hat{\phi}_{p,i}))^T COV_{3 \times 3,i}^{-1}(\hat{\phi}_{p,i}) (\vec{r} - \vec{r}_i(\hat{\phi}_{p,i})). \quad (5.5)$$

Since we already have the tracks in the perigee representation, we can consider a transformation of variables to write the $\chi^2(\vec{r})$ in the perigee representation.

$$\chi^2(\vec{r}) = \sum_{i=1}^N (\vec{q}_{pred,i}(\vec{r}) - \vec{q}_{meas,i})^T COV_{5 \times 5,i}^{-1}(\vec{q}_{pred,i}(\vec{r}) - \vec{q}_{meas,i}), \quad (5.6)$$

where the $\vec{q}_{pred,i}(\vec{r})$ are the track parameters if the track actually passed through the vertex location \vec{r} . The projection from 3-dimensions to 5-dimensions is undefined, so the track momenta at the vertex location are also added as inputs to the fit so that the fit becomes defined.²

Two *fast* ways to find the \vec{r} that minimizes this $\chi^2(\vec{r})$ are the Billoir method [72] and using the kalman filter [66]. In the Kalman filter approach, tracks parameters are iteratively added to the vertex fit, and the “state vector” includes our best knowledge of the vertex location.

Adaptive vertex fitting

The vertex fit described above will always return a fit under the hypothesis that all of the input tracks originated from the same vertex. However, the result of vertex fit can be influenced by *outliers*, which could be poorly reconstructed (or fake) tracks, or good quality tracks that originated from another location which could pulled the vertex fit. Since the full χ^2 is the sum of the contributions

²This works b/c the perigee representation is the track momenta + a spatial location.

of the individual tracks, we can look at the χ^2 of an individual tracks to determine if it is an outlier using $\text{Prob}(\chi^2) = 1 - \int_0^{\chi^2} f(\tilde{\chi}^2) d\tilde{\chi}^2$. Outlier tracks are removed and the vertex fit is repeated.

Adaptive vertex fitting methods assign a weight to each track in the vertex fit to integrate the track selection into the vertex fit. This weight is given by the $\omega(\hat{\chi}_i^2)$, updating the vertex $\chi^2(\vec{r})$ below:

$$\chi^2(\vec{r}) = \sum_{i=1}^N \omega(\hat{\chi}_i^2) (\vec{q}_{pred,i}(\vec{r}) - \vec{q}_{meas,i})^T COV_{5 \times 5,i}^{-1} (\vec{q}_{pred,i}(\vec{r}) - \vec{q}_{meas,i}). \quad (5.7)$$

To allow the fit to progressively update the weights during the fitting procedure, we use a temperature-dependent Fermi function:

$$\omega(\hat{\chi}^2) = \frac{1}{1 + \exp\left(\frac{\hat{\chi}^2 - \chi_{cutoff}^2}{2T}\right)}, \quad (5.8)$$

Figure 5.6(a) shows this function for a few representative values of T . As $T \rightarrow \infty$, this function is uninformative and flat (just 0.5). As T decreases, this function gets an increasingly sharp turn-off, where $\omega(\hat{\chi}^2) = 1$ includes the track is included in the fit, and $\omega(\hat{\chi}^2) = 0$ removes it.

At the beginning of the fit, a high temperature gives a low certainty for all the track weights, and then an annealing schedule will make the weights increasingly more deterministic as the temperature is decreased. Figure 5.6(b) shows the track weights at the beginning ($T=1$), middle ($T=4$) and end ($T=64$) of the fit annealing procedure. At the end, tracks below a χ_{cutoff}^2 of nine are removed from the fit.³

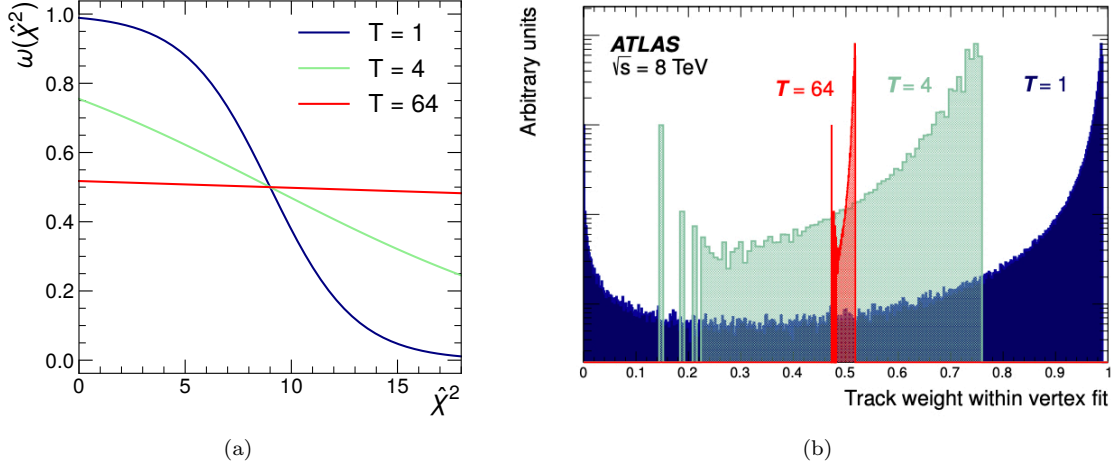


Figure 5.6: **Left:** Fermi function (Eq. 5.8) for a range of different temperatures, $T = 1, 4, 64$. **Right:** Corresponding track weights for the same three temperatures [73].

³ $\chi_{cutoff}^2 = 9$ is the χ^2 value corresponding to approximately three standard deviations.

5.2.2 Primary vertex reconstruction

The Run 2 primary vertex reconstruction algorithm uses an *iterative vertex finder*, with the steps outlined below [73]:

1. Seed the vertex position using the center of the beam spot for the x, y position, and the half mode sampling algorithm [74] to find regions with a high density of tracks along the z axis.
2. Then the vertex is fit using an adaptive vertex fit [75]. The annealing temperature schedule starts at $T = 64$ and converges at $T = 1$.
3. The final weights for the tracks are used in the final vertex fit, only tracks with a χ^2 values incompatible by more than seven standard deviations are considered as “outliers” and removed from this vertex fit.
4. To find the rest of the vertices, steps 1 – 3 are repeated to create vertices until all of the tracks have been associated to vertices.

Figure 5.7 shows the result of this vertex reconstruction. Of the vertices with at least two tracks, the **primary vertex** (PV) is defined as the reconstructed primary vertex with largest $\sum p_T^2$ of the associated tracks. After the primary vertex is found, the track parameters are refit including the PV as an additional constraint for the fit. However, for b -tagging (described in Chapter 6), we use tracks that don’t include this PV constraint.

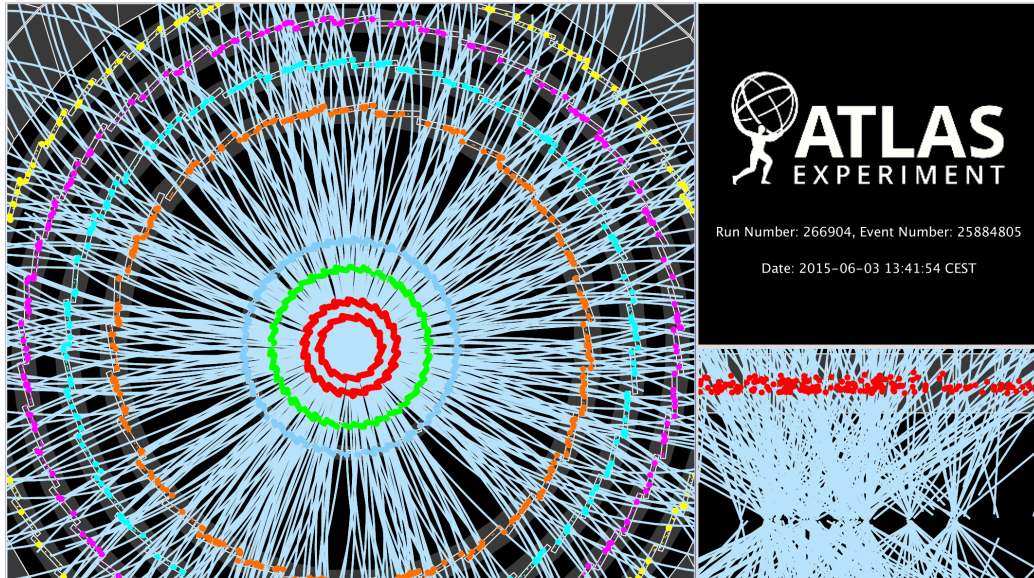


Figure 5.7: Event display from one of the first Run 2 events. In the left image shows the (r, ϕ) view, and the colored dots show the hits in the Silicon detectors (IBL, pixel and SCT). In the lower right image, the z axis is oriented horizontally, and there are set of reconstructed vertices from a single bunch crossing [76].

5.3 Jets

5.3.1 Anti- k_t jet clustering algorithm

Quarks and gluons produce a collimated spray of particles in the detector, and a *jet* is a proxy for the kinematics of the original quark or gluon. A clustering algorithm groups together constituents in the event (which could be calorimeter energy deposits, tracks, or particle flow objects) to define the jets. Infrared and collinear (IRC) safety is a nice property for a jet definition, where an infrared-safe algorithm is robust to additional soft particles which have $E \rightarrow 0$, and a collinear-safe algorithm is robust for a single particle with energy E splitting its energy into two particles in the same direction. Sequential jet clustering algorithms are a set of IRC-safe algorithms which cluster jets based on distance metrics:

$$d_{ij} = \min \left(p_{Ti}^{2p}, p_{Tj}^{2p} \right) \frac{\Delta R_{ij}}{R^2} \quad d_i = p_{Ti}^{2p}. \quad (5.9)$$

where $\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, R is the jet-radius, and p balances the energy and geometric scales of the event. $p = -1$ defines the anti- k_t algorithm [77] which is now the standard jet clustering algorithm in collider physics. Other choices of p correspond to other jet clustering algorithms considered in the past: $p = 1$ is the k_t algorithm which clusters the jets by undoing the fragmentation process, and $p = 0$ is the Cambridge/Aachen algorithm which just uses the angular distances to cluster the jets. The clustering algorithm proceeds by the following steps:

- The d_{ij} and d_i are calculated for all particles in the event. Then $d \equiv \min(d_{ij}, d_i)$.
 - If $d = d_{ij}$, jet i and jet j are combined.
 - If $d = d_i$, then jet i becomes a final jet.
- The previous steps are repeated until all particles in the event are been clustered.

The anti- k_t algorithm starts by clustering the radiation around the hardest particles in the event since the leading p_T particle will define the $\min \left(\frac{1}{p_{Ti}^2}, \frac{1}{p_{Tj}^2} \right)$ term in the d_{ij} definition. This allows jets in the event have a stable direction early on in the combination. The anti- k_t algorithm is preferred to other sequential jet algorithms since jets have regular boundaries which are approximately conical, as shown in Figure 5.8. It's also a “greedy” algorithm since if two jets have a distance $R < \Delta R(j_1, j_2) < 2R$, the higher p_T jet will be conical while the subleading jet will have a semi-circle shape, as seen by the green and pink jets in Figure 5.8.

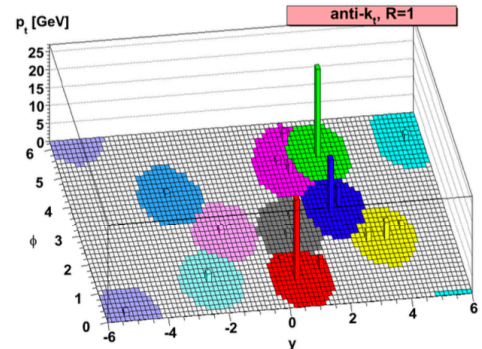


Figure 5.8: Illustration of the shapes jets created by the anti- k_t algorithm [77].

5.3.2 Calorimeter Jets

One way to reconstruct jets is based on energy deposits in the calorimeter. A 3d reconstruction algorithm first clusters the signals from the calorimeter cells into *topo-clusters*, which approximates the energy deposits of individual hadrons [61]. For each calorimeter cell, a “significance” is defined:

$$\zeta_i = \frac{E_i}{\sigma_{noise,i}} \quad (5.10)$$

where σ_{noise} includes both the contributions to the noise from the electronics noise and pile-up. The clustering is done with the prescription below.

1. Topo-clusters are “seeded” with cells that have $|\zeta_i| > 4$.
2. Then the topo-clusters are grown by including the cells adjacent to the seed cells with $\zeta_i > 2$.
3. Finally, all adjacent cells are added to the topo-cluster.

Then there’s a splitting step where a topo-cluster is allowed to split into two if there are two local energy maxima inside of the topo-cluster.

The cells are clustered with $|\zeta_i|$ as the calorimeter read-out signals are *shaped* so that the total integral of the signal is 0 (see Figure 4.14). Negative energy contributions in the topo-cluster reconstruction thus cancel the contributions of in-time PU and out-of-time PU. The jets clustered from topo-clusters with a distance parameter of $R = 0.4$ are called *EMTopo jets*, and are the proxies for the individual quarks and gluons. In the jet reconstruction, only the topo-clusters with positive net energy are included.

5.3.3 PFlow Jets

As Eq. 4.5 and Eq. 4.8 emphasize, the resolution the tracker is better for low p_T , while the resolution of the charged energy components improves at high p_T . To maximally benefit from the strengths of the separate subsystems, the tracks can be used to estimate charged hadron energy in the jet [62]. To avoid double counting the energy from the tracks and clusters, a charged energy subtraction scheme is needed to remove the associated charged hadron clusters. A schematic of what this process looks like is shown in Figure 5.9.

Tracks are matched to topo-clusters using proximity in (η, ϕ) space, also accounting for the size of the topo-cluster. The tracks are only “matched” if the cluster carries more than 10% of the track’s momentum. Sometimes the topo-cluster fails to cluster all of the hadron’s energy in a single topo-cluster. In cases where the expected energy of the track is less than the expected track’s energy, a “split shower recovery” combines nearby topo-clusters to form a topo-cluster set. (This step has no impact in Figure 5.9 as both of the hadrons are already reconstructed as separate topo-clusters.) From this topo-cluster set, the expected energy of the track is subtracted from the topo-cluster’s

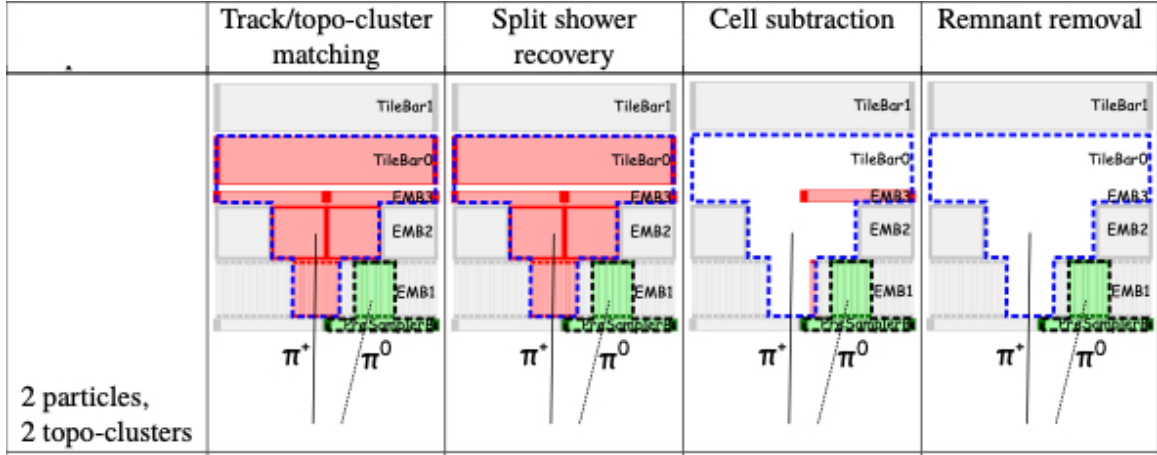


Figure 5.9: Illustration of the charged energy subtraction algorithm for reconstructing the PFlow objects for two hadrons π^+ and π^- with have been reconstructed in two topo-clusters (shown in the dashed blue and black lines). The pre-sampler, three ECAL layers, and first two TileCal layers are shown, and the energy of the π^+ is in red and the energy of the π^0 is in green [62].

cells, starting with high-energy density cells. If the residual energy is consistent with the resolution of the expected track energy, the residual energy is also subtracted in the last step called “remnant removal”.

The result of this algorithm is a set of tracks, modified and unmodified topo-clusters which are the particle-flow (PFlow) objects. The PFlow objects can also clustered with the anti- k_t algorithm and the same $R = 0.4$ to form *PFlow jets*.

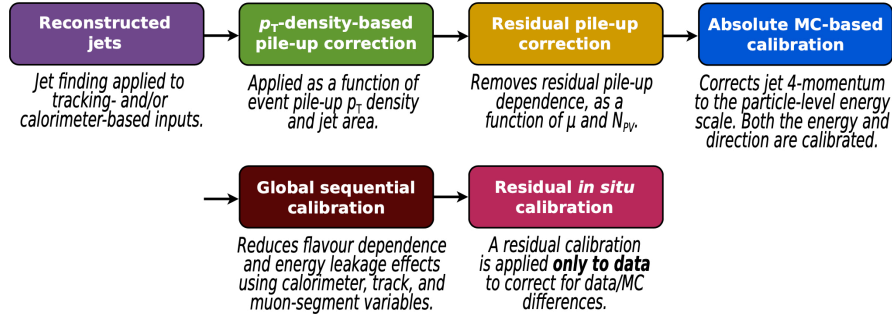


Figure 5.10: Steps of the 4-momentum calibration for PFlow jets [78].

Figure 5.10 enumerates the steps by which the 4-momentum of these reconstructed jets is corrected to match the kinematics of a truth jet⁴. The first three corrections account for contamination from the underlying pile-up distribution and fluctuations due to the origin of the jet [78]. The the

⁴The truth jets come from the anti- k_t clustering of the stable final state truth particles (hadrons and charged leptons) in simulation.

Global Sequential Calibration improves the jets p_T resolution (and associated uncertainties) by sequentially removing the dependence of the reconstructed jet response ($R = E^{reco}/E^{true}$) on key event observables. Finally, the residual differences between data and MC are accounted for by measuring the momentum imbalance in Z +jets, γ +jets and multi-jet events [78].

The corresponding resolution on the jets p_T and ϕ is compared in Figure 5.11. The better p_T resolution of the tracker at low p_T translates into a better resolution for low p_T jets in Figure 5.11(a). Additionally, the better angular resolution of the tracker also translates into a better angular resolution on the jet, in Figure 5.11(b) shown for the jet ϕ – which again has a larger impact at low jet p_T .

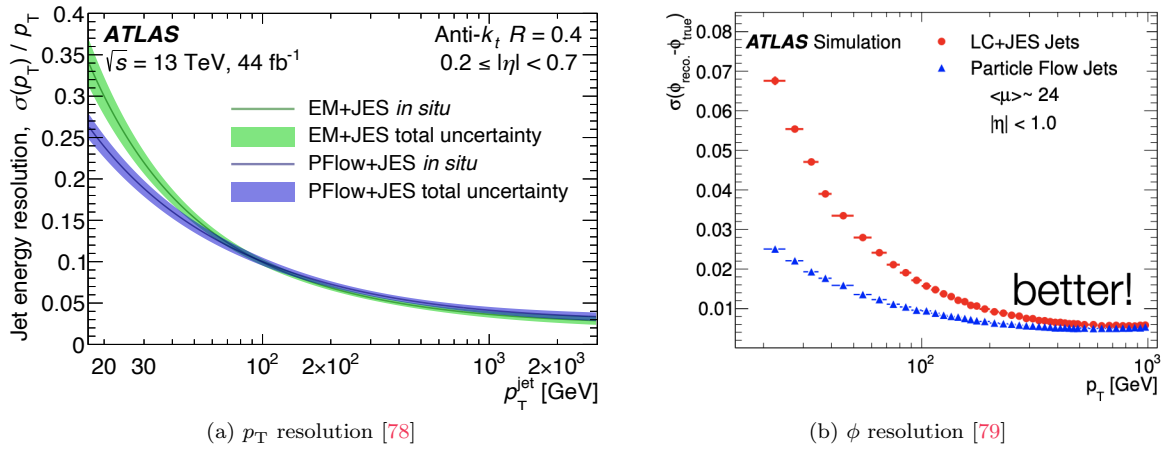


Figure 5.11: Improvement due to the PFlow jet reconstruction algorithm.

5.4 Pileup jet rejection

We are interested in the jets coming from the quarks and gluons in the hard scatter (HS) process, which will be called HS jets. As mentioned in Chapter 3, in Run 2, there are on average 34 collisions per bunch crossing, and we want to identify the HS jets in the midst of the jets that originate from the background pile-up (PU) vertices, called PU jets.

There are two types of PU jets we try to reject. The first are jets from quarks and gluons that originate from a vertex other than the primary vertex, as depicted by the red jet in Figure 5.12. The second type of PU jet can be formed from clustering hadrons that originated from several different vertices, as illustrated by the orange jet in Figure 5.12. This type of PU jet is also called a “stochastic jet”.

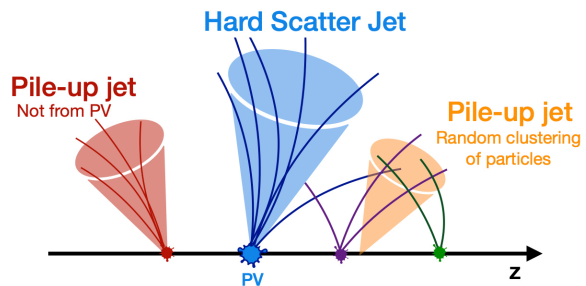


Figure 5.12: Schematic of a hard scatter jet (in blue) and the two types of pile-up jets that we try to reject (in red and orange).

The Jet Vertex Fraction (JVF) probes whether a jet came from the primary vertex:

$$\text{JVF} = \frac{\sum_k p_T^{\text{trk}_k}(\text{PV})}{\sum_k p_T^{\text{trk}_k} + p_T^{\text{PU}}}.$$

The numerator, $\sum_k p_T^{\text{trk}_k}(\text{PV})$, is the scalar sum p_T of the tracks in the PV, while the denominator is the scalar sum p_T of the tracks in the PV and PU vertices, as the second term is defined as $p_T^{\text{PU}} = \sum_{n|\text{vtx}_n \neq \text{PV}} \sum_l p_T^{\text{trk}_l}(\text{vtx}_n)$. Since p_T^{PU} increases with the number of pile-up vertices in the bunch crossing, we use a corrected JVF (corrJVF) to ameliorate the PU dependence by dividing p_T^{PU} by the number of PU tracks, $n_{\text{trk}}^{\text{PU}}$:

$$\text{corrJVF} = \frac{\sum_k p_T^{\text{trk}_k}(\text{PV})}{\sum_k p_T^{\text{trk}_k} + \frac{p_T^{\text{PU}}}{\alpha \cdot n_{\text{trk}}^{\text{PU}}}},$$

where $\alpha = 0.01$, which is the slope of $\langle p_T^{\text{PU}} \rangle$ versus $n_{\text{trk}}^{\text{PU}}$ [80].

To help reject stochastic jets, the variable R_{p_T} measures the charged energy of the jet that originated from the PV:

$$R_{p_T} = \frac{\sum_k p_T^{\text{trk}_k}(\text{PV})}{p_T^{\text{jet}}}.$$

PU jets have an R_{p_T} peaked at zero and steeply falling, while HS jets have higher R_{p_T} values. R_{p_T} only considers the tracks from the PV, and is independent of the number of PU vertices to first order.

These two variables, corrJVT and R_{p_T} , are used together to define the Jet Vertex Tagger (JVT). corrJVT and R_{p_T} are rescaled so they have the same range, and a kNN algorithm uses the 100 nearest neighbors (by the Euclidean metric) approximates the likelihood that the jet is from a HS, which defines the JVT discriminant [81]. As the jets that don't originate from the HS are generally softer, the JVT cut is only applied for jets with $p_T < 60$ GeV and $|\eta| < 2.4$. The default JVT working point is 96% efficient for hard scatter jets.

5.4.1 Boosted jets

Jets clustered from calorimeter clusters with a distance parameter of $R = 1.0$ are used as a proxy for boosted Higgs, W, and Z bosons. Large-R jets are clustered with topo-clusters calibrated to the hadronic energy scale.

The contribution from pile-up collisions, initial state radiation, and the underlying event gives an approximately uniform distribution of energy in the calorimeter, and a “trimming” procedure is used to mitigate the impact of this uniform energy inside of the large-R jet [82]. The constituents of the large R jet are reclustered into “sub-jets” with a distance parameter of $R = 0.2$. Sub-jets with a p_T less than 5% of the large R jet p_T are removed from the large-R jet [83].

Higgs and vector bosons decay into two particles which give two hard prongs inside of the large-R

jet. Searching for the substructure of the number of prongs inside of the large- R jet has been a sub-field of intense study in recent years (see [23] for a recent review). To identify the hard prongs in the large- R jet, ATLAS uses jets clustered from the tracks, called “track jets” as illustrated in Figure 5.13 [83].

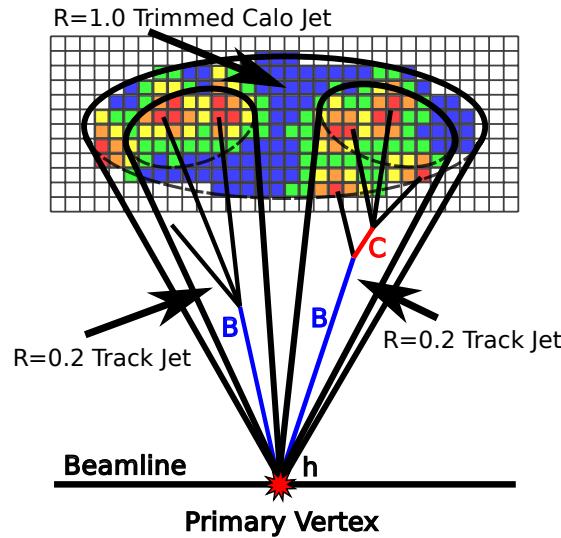


Figure 5.13: Illustration of the large- R jet defined from the calorimeter clusters, and the track-jets which identify the two hard prongs inside the jet. They provide a jet axis for b -tagging algorithms [83].

5.4.2 Variable Radius track jets

Additionally, we can form jets using the anti- k_t algorithm to cluster tracks, which are called *track jets*. These are useful for identifying boosted Higgs bosons based on the contained sub-jets inside. Running b -tagging on track jets presents a nice application, as tracks provide the most relevant information for b -tagging. In early Run 2, track jets clustered with an $R = 0.2$ distance parameter were used for identifying boosted Higgs bosons [84]. However, as we produce Higgs bosons that are even more energetic, the decay products become increasingly more collimated until the b -hadrons are no longer reconstructed into separate subjets, as illustrated in the left image Figure 5.14.

A modification to the anti- k_t algorithm that allows for the flexibility to account for a finer track jet resolution as the b -hadrons become more collimated uses Variable Radius (VR) for reconstructing the jets [85]. The distance parameter of the anti- k_t is inversely proportional to the p_T of the proposed jet:

$$R_{eff}(p_T) = \frac{\rho}{p_T}, \quad (5.11)$$

where ρ is a parameter that sets the scale of the jet radius. Additionally, to avoid the reclustered jets from becoming unphysically large or small, minimum and maximum radii, R_{min} and R_{max} , bound

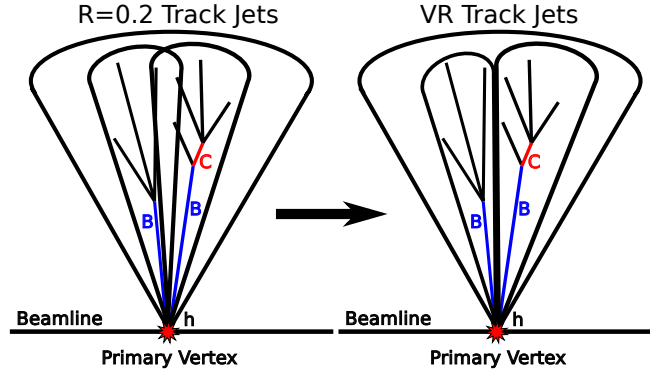


Figure 5.14: Illustration of the motivation for VR track jets. The left depicts that with fixed radius track jets, for high enough boosts the products become so collimated that separate small- R jets cannot be reconstructed. The right figure illustrates how the decreasing radius with increasing sub-jet momentum can allow the separate sub-jets reconstruction of the individual b -hadron decays [83].

the range of R_{eff} . The parameters that ATLAS uses for the VR track jet collection are $\rho = 30\text{ GeV}$, $R_{min} = 0.02$ and $R_{max} = 0.02$, which were optimized to maximize the probability that reconstructed sub-jets are within $\Delta R < 0.3$ of the b -hadrons.

Figure 5.15 illustrates that the $R=0.2$ fixed radius track jets (in the red curve) have a very poor efficiency for matching to both of the b -hadrons as the Higgs jet gets boosted beyond a TeV. The VR track jet alternatives are able to keep a mostly efficiency at these higher boosts. The blue line of $\rho = 30\text{ GeV}$ is the parameter that was adopted by ATLAS. Track jets and variables radius track jets are relevant to my contributions to b -tagging and will be discussed in Chapter 6.4.2.

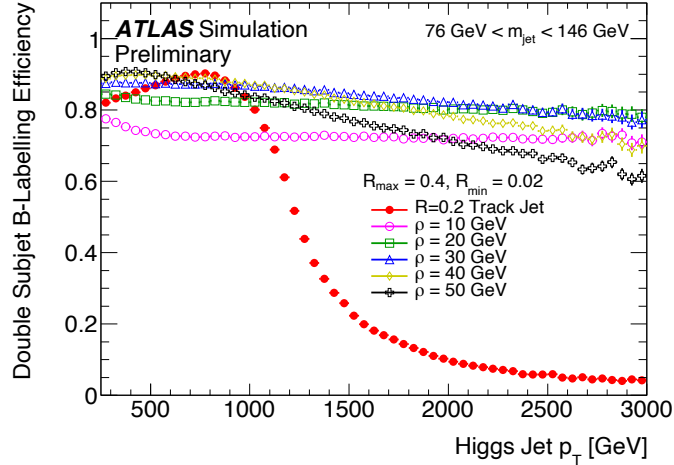


Figure 5.15: Comparison of the truth labelling efficiency for the fixed $R=0.2$ track jet, and the VR radius track jets with different choices of ρ . The truth labelling efficiency is the probability that each of the b -hadrons is within $\Delta R = 0.3$ of the sub-jet axis [83].

5.5 Leptons and Photons

The reconstructed leptons: electrons, muons, and τ s play a less central role in this thesis, and are described briefly in this section. Since photons have a nearly identical signature to electrons, their reconstruction is described simultaneously in the following subsection.

5.5.1 Electrons and photons

As illustrated in Figure 4.2, both electrons and photons deposit most of their energy in the calorimeter. The energy of the electron or photon is defined by a corresponding “super-cluster” [86]. The topo-clusters (described in Chapter 5.3.2) are defined similarly here. The topo-clusters considered for electron and photon reconstruction are only considered further if more than half of the energy of the topo-cluster is in the EM calorimeter. This set of topo-clusters are used as “seed clusters”, and close-by satellite clusters are added to naturally account for the larger cells from Bremsstrahlung photons or the electrons from photon conversions. The combined seed and satellite clusters form the super-cluster. If the super-cluster matches to a track, then the object is an electron. Photons are reconstructed either from no corresponding electron tracks, or if the super-cluster is matched to an e^+e^- conversion vertex. For cases that could be ambiguous, the shower shape information and the associated track(s) is included to distinguish between electrons and photons [86].

5.5.2 Muons

Muons are minimum ionizing particles, so they make it to the outer (appropriately named) muon spectrometer. This muon system measuring the trajectory far away from the interaction point allows for a large lever arm to improve the momentum resolution.

The muon spectrometer (MS) tracks are reconstructed by searching for consecutive hits that lie along the same line in a single muon station. The tracks segments between muon stations are combined, now taking into account the curvature from the muon spectrometer’s magnetic field. The MS track is then combined with an inner detector (ID) track to get a global χ^2 fit of the track’s parameters, accounting for the energy lost by the muon in the calorimeter. After this global χ^2 fit, hits from outliers are removed and the fit is redone.

The “Medium” (default) muon identification working point also reconstructs muons from ID tracks extrapolated to the muon spectrometer that have at least three consecutive hits in the first muon station. This improves the efficiency for low p_T muons that stop before reaching the second muon station. In $t\bar{t}$ events, 98% of the “Medium” muons come from combined muon with a combined ID and MS track [87].

5.5.3 Taus

A τ lepton has a mass of 1.777 GeV, a proper decay length $87\mu\text{m}$, and can decay either leptonically or hadronically. The leptonic tau decays have $\tau \rightarrow \nu_\tau \nu_\ell \ell$, where $\ell = e, \mu$, and the τ is inferred from the electron or muon in the final state.

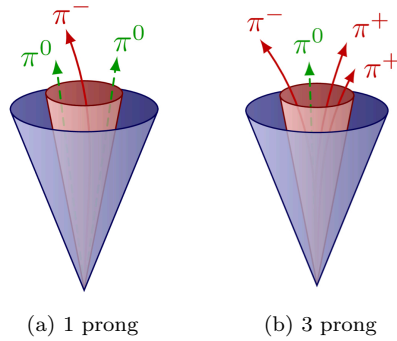


Figure 5.16: Examples of hadronic τ -decays [88].

In Run 1 and early Run 2, a BDT was used to discriminate between hadronic τ decays from other jets, but In Run 2, a recurrent neural network (RNN) hadronic τ classification algorithm was employed [89]. Since the tau-tagger closely follows the b -tagger, we will hold the discussion of RNNs for particle identification until Chapter 6.3.2. Figure 5.17 shows the general architecture for τ s. Two RNNs process the tracks and the topo-clusters in the jet, where the topo-clusters include information about the photons from the $\pi^0 \rightarrow \gamma\gamma$ decays into the τ jet. Different NNs are trained for one and three-prong τ decays, with slight differences in the high-level variable set to account for the uniqueness of the two topologies. This switch from the BDT to the RNN-based classifier improved the jet misidentification rate by 50% – 75% [89].

The majority (65%) of the τ decays are in the hadronic decay mode, with most of these decays with either one or three charged pions in the final state. To reconstruct the τ decay topology, a BDT classifies the tracks in the jet into core and isolation tracks, where the number of core tracks the jet defines the number of “prongs” or charged pions in the tau decay. Figure 5.16 shows examples of these one and three-prong τ decays. There can also be a number of neutral pions in the final state.

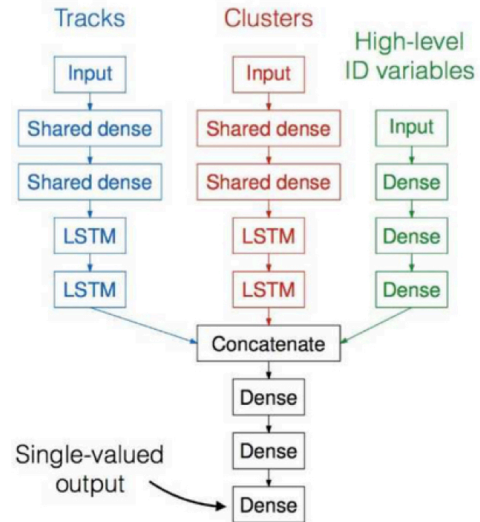


Figure 5.17: Set up for the hadronic τ identification. An LSTM is a type of RNN architecture.

5.6 E_T^{miss}

Since the partons collide head-on, the momentum is conserved in the transverse plane. The presence of particles that don't interact in the detector (such as neutrinos in the SM) can be inferred by a momentum imbalance. This is quantified by the negative vectorial sum over the calibrated physics objects and the rest of the tracks in the hard scatter (HS) vertex [90]:

$$\vec{E}_T^{miss} = - \sum_{i \in \{e, \gamma, \tau, \mu, \text{jets}\}} \vec{p}_T^{(i)} - \sum_{\text{unused tracks}} \vec{p}_T^{\text{track}}. \quad (5.12)$$

The first term corresponds to the hard objects in the event - and an overlap removal scheme ensures the included objects don't double count calorimeter energy deposits. The last term includes the transverse momentum for the tracks in the PV that were not included in the reconstruction of the other hard objects. Only including the HS tracks in this vertex helps suppress the impact of PU in the \vec{E}_T^{miss} calculation. Then $E_T^{miss} = |\vec{E}_T^{miss}|$, and is by definition non-negative. Large E_T^{miss} values are expected from neutrinos or other weakly interacting BSM particles, but non-zero E_T^{miss} values can also arise from miscalibrated physics objects, the limited acceptance for the physics objects, or detector noise. The resolution on E_T^{miss} is between 10% – 25% for $t\bar{t}$ events [90].

Part II

Flavor tagging

Here we introduce the algorithms for identifying the jets from b -quarks. Chapter 6 reviews the algorithms used for the Run 2 b -taggers, with an emphasis on the RNNIP optimization, which is a contribution of this thesis work. Chapter 7 describes a novel tagger, DIPS, designed to replace RNNIP, and we present further input optimizations which improved the performance of this track-based b -tagger.

6

ATLAS Run 2 b -taggers

Technical expertise is the mastery of complexity – creativity is the mastery of simplicity.

– Michael Nielson, *Neural Networks and Deep Learning*

For studies of the Higgs boson properties, identifying the jets initiated by b -quarks (or, b -tagging) is a fundamental tool. The b -hadron decay to a c -hadron is mediated by the weak force with a characteristically “long” lifetime of 1.5 ps, which is long enough to reconstruct the secondary vertex, as shown in Figure 6.1. We seek to identify b -jets in presence of two types of background jets. The first is light jets (or jets initiated by gluons, up, down, or strange quarks) that have most of the tracks originating from the primary vertex (PV), which are called “prompt tracks”. The second background is c -jets which also have a weakly decaying c -hadron with a 0.6 ps lifetime, and are the more challenging background to reject.

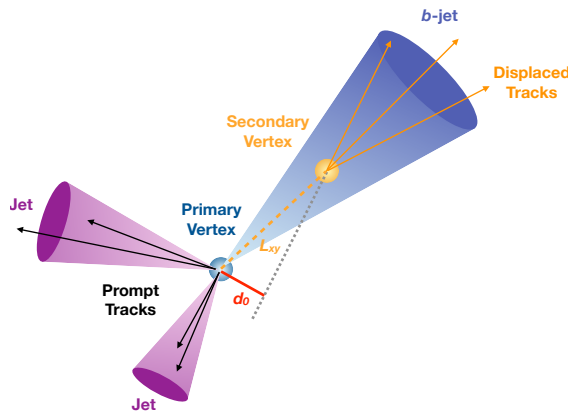


Figure 6.1: Schematic for the characteristic “long” lifetime of a b -hadron [91].

In addition to the long lifetime of the b -hadron, we have a few other handles for b -jet identification.

The fragmentation function for the b -quark gives the b -hadron $\approx 70\%$ of the initial b -quark's momentum. Thus, the displaced tracks from the heavy flavor decay are more energetic than the fragmentation tracks in light jets. The b -quark has a large mass of ≈ 5 GeV, so the reconstructed displaced vertex gives a high-mass hadron. Figure 6.2 shows such a displaced b -hadron decay (with the cyan tracks) in the ATLAS detector, and the reconstructed vertex mass is 2.5 GeV. b -jets have a large number of displaced tracks from the heavy flavor decay, with the b -hadron decaying on average into five charged particles. These tracks from the long-lived hadron decay are displaced from the PV have a large impact parameter (IP) when the track is extrapolated back to the point of closest approach to the PV. Information about the quality of the tracks helps determine if the tracks' displaced IP is trustworthy or due to mismeasured tracks that actually originated from the PV. Figure 6.2 indicates the hits and holes on one of the displaced tracks, and information about the hit patterns in the inner layers is used by some of the b -tagging algorithms.

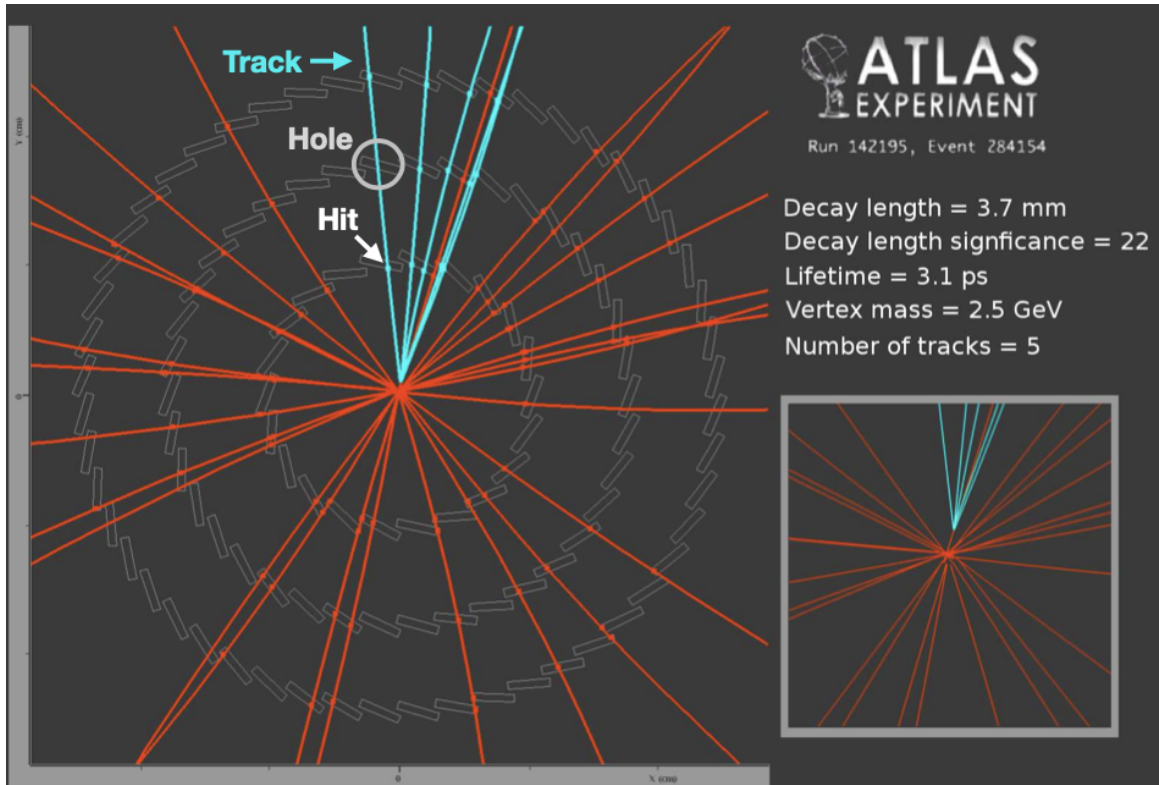


Figure 6.2: Illustration of what a b -decay looks like in the ATLAS detector. The cyan colored lines illustrate the tracks from the b -hadron decay, and in the inset figure shows the displacement of these tracks from the primary vertex. Only three pixel layers are shown as this is a Run 1 event, and the IBL was not yet installed [92].

Light jets can fake a b -jet signal either through mismeasured tracks causing the track to have a large IP or “fake tracks” formed from the random combination of hits seeming to have a large IP. c -jets are the more challenging to reject, because, in addition to being long-lived, the c -hadron also has a heavy mass of 1.8 GeV. The c -quark additionally also has a hard fragmentation function giving high p_T tracks from the displaced decay.

This chapter is organized as follows: Chapter 6.1 describes the datasets and selections used to train and evaluate the algorithms. Chapter 6.2 details how we will evaluate the performance. Chapter 6.3 revises the low-level algorithms which rely on the displaced tracks or reconstructed displaced vertices, including the recurrent neural network (RNN) optimizations central to this thesis work. Chapter 6.4 introduces the high-level DL1r classifier which uses the outputs from the low-level algorithms. Finally, Chapter 6.5 summarizes with the calibration of these taggers.

6.1 Datasets

6.1.1 Hybrid sample creation

PFlow jets

ATLAS recently switched from using EMTopo jets (which reconstruct a jet from the topo-clusters in the calorimeter) to using PFlow jets (which cluster a jet using particle flow candidates to take advantage of the better resolution of the tracker for lower momenta objects). This improves the jet reconstruction by improving both the jet p_T and angular resolution (shown by Figure 5.11). The latter is crucial for b -tagging as the lifetime signing of the tracks uses the jet axis to approximate the b -hadron flight direction (more details in Chapter 6.3.1). This effort of optimizing the b -taggers on the new PFlow jets collection also included the first time the RNN training was recommended to the collaboration.

For these newly optimized taggers to be maximally useful for ATLAS’s diverse physics program, the training sample was designed to cover a large p_T range for b , c , and light jets. To get a b -jet enriched physics sample to train the taggers and evaluate the performance, $t\bar{t}$ is often used since the large CKM element of $V_{tb} = 0.999$ means the top almost always decays to a b -quark. The $t\bar{t}$ training statistics are limited for jets p_T above ≈ 300 GeV, so the $t\bar{t}$ MC training sample was combined with a BSM Z' sample to increase our coverage for very high p_T jets [93]. To validate the robustness of these trainings, two different Z' samples were considered:

- **Standard Z' sample:** Jets are produced up to 1.5 TeV
- **Extended Z' sample:** Jets are produced up to 5 TeV.

The Z' sample is combined with $t\bar{t}$ to produce a “hybrid sample”. The b -jets of this sample are split based on the b -hadron p_T where jets with a b -hadron $p_T < 250$ GeV are taken from the $t\bar{t}$ sample

and otherwise b -jets are taken from the Z' sample. The c and light-jets are sampled from the $t\bar{t}$ and Z' samples with a 250 GeV jet p_T cut. Two hybrid samples were considered: a **hybrid 1.5 TeV** sample built from the 1.5 TeV Z' sample and another **hybrid 5 TeV** sample built from the 5 TeV Z' sample. Our recommendation uses the hybrid 5 TeV training sample, with the resulting p_T spectrum shown in Figure 6.3. The cut-off at 250 GeV due to the trade-off between the two samples is clearly visible, and the maximum p_T starts falling off beyond 5 TeV p_T jets – close to the maximum jet p_T that could be produced by a heavy resonance decay in 13 TeV pp collisions.

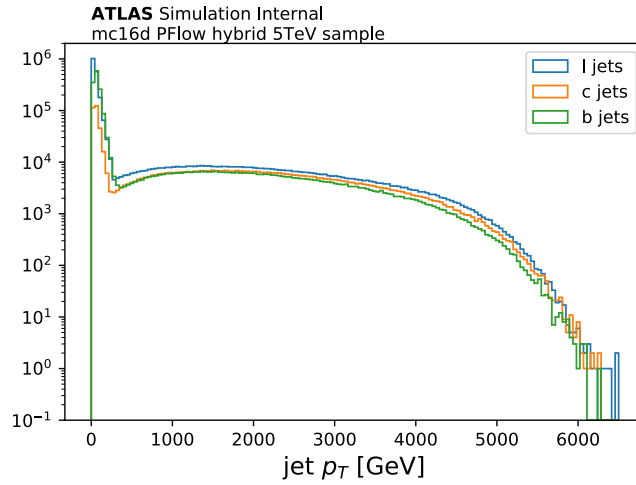


Figure 6.3: The p_T spectrum for training the Run 2 b -tagger recommendations (RNNIP and DL1r).

VR track jets

Additionally, we provided the first dedicated track jet retraining recommendations for the b -taggers. Physics analyses with track jets is a recent innovation, with [84] detailing the studies of boosted Higgs tagging applying b -tagging for fixed radius track jets. Here the b -taggers are optimized for the variable radius jet collection (described earlier in Chapter 5.4.2). Further details on the VR hybrid sample are in Chapter A.1.

6.1.2 Sample details

The simulated $t\bar{t}$ events are for $\sqrt{s} = 13$ TeV proton-proton collisions, and at least one of the W bosons from the tops decays leptonically. Events are generated at next-to-leading order using the POWHEGBOX [94–97] v2 generator and interfaced to PYTHIA 8.230 [98] to model the parton shower, hadronization, and underlying event, with parameters set according to the A14 tune [99]. The decays of b - and c -hadrons are performed by EVTGEN v1.6.0 [100]. Particles are passed through the ATLAS detector simulation [101] based on GEANT4 [102].

In the sample generation, the Z' is set to decay to b , c , and light quarks with equal probability. The matrix element for the Z' propagator was additionally modified to produce a sample with a flat p_T spectrum.

6.1.3 Jet selection

PFlow jets have $p_T \geq 20$ GeV, $|\eta| < 2.5$, and are required not to overlap with a generator-level electron or muon from W boson decays. Additionally, the contamination of jets from other interactions in the beam crossing (pile-up) is suppressed by applying the jet vertex tagger optimized for particle flow jets (see Chapter 5.4). Jets are labeled as b -jets if they are matched to at least one b -hadron having $p_T \geq 5$ GeV within $\Delta R(b\text{-hadron, jet}) < 0.3$ of the jet axis. If this condition is not satisfied, then c -hadrons and then τ leptons are searched for, with similar selection criteria. If a jet is matched to a c -hadron (τ -lepton), it is labeled a c -jet (τ -jet). A jet that does not meet any of these conditions is called a light-jet.

The track features are crucial for b -tagging, and tracks are associated to jets using a ΔR association cone that is a function of jet p_T (in GeV):

$$\Delta R(\text{track, jet}) < 0.239 + \exp \left[-1.220 - 1.64 \times 10^{-2} \cdot p_T \right]. \quad (6.1)$$

The maximum association $\Delta R(\text{track, jet})$ is approximately 0.45 for a jet with $p_T = 20$ GeV, and $\Delta R(\text{track, jet})$ is at 0.25 (close to the plateau) when the jet $p_T = 200$ GeV. If a track is within the association cones of more than one jet, it is assigned to the jet which has a smaller $\Delta R(\text{track, jet})$.

6.2 Evaluating tagger performance

A multi-class classifier with three output nodes for the b , c , and light jet probabilities (denoted p_b , p_c , and p_l , respectively) can be combined into a single b -tagging discriminant:

$$D_b = \log \frac{p_b}{(1 - f_c)p_l + f_c p_c}, \quad (6.2)$$

where f_c is a parameter that balances the light versus c -jet rejection, and is chosen after training the tagger. An example D_b distribution for each jet flavor is shown in Figure 6.4. Clear separation between the distribution of b -jets and light jets can be seen, as well as a strong but smaller separation between b -jet and c -jets because of the similarities between b -hadron and c -hadron decays.

The performance of taggers can be examined and compared through a Receiver Operator Characteristic (ROC) curve: a scan is performed for a discriminant threshold d on D_b , and the *efficiency* for b -jets at each threshold is computed as the fraction of b -jets with $D_b > d$, while the *rejection* of c -jets and light jets is computed as one over the mistag efficiency for c -jets or light jets,

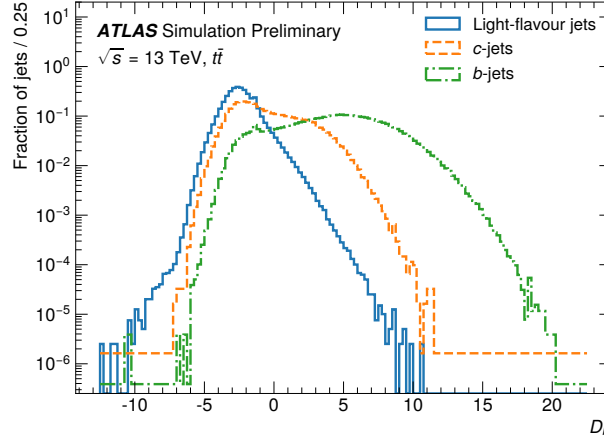


Figure 6.4: Distributions of a b -tagging discriminant as defined in Eq. 6.2 with $f_c=0.07$, for b , c , and light-jets [5].

respectively. The b -jet efficiency and light (or c) jet rejection for the same d are then plotted.

A working point (WP) is defined with the b -jet efficiency in a $t\bar{t}$ sample, and the ATLAS working points are 60%, 70%, 77%, and 85%. The calibration for each of the jet flavors are done with pseudo-continuous (PC) bins defined from these working points, as given in Table 6.1.

Pseudo-continuous bin	b -tagging efficiency
1	$> 85\%$
2	$77\% - 85\%$
3	$77\% - 85\%$
4	$77\% - 85\%$
5	$< 60\%$

Table 6.1: Pseudo-continuous (PC) working points definitions. PC bin 5 is the highest b -jet purity bin, and PC bin 1 fails all of the b -tagging WPs.

While ROC curves give a global view of an algorithm’s performance, the behavior of the b -tagging efficiency and the background rejection as a function of key kinematic variables is also vital for understanding the performance within analyses. To explore this metric, after a “fixed WP” threshold is determined, the b -jet efficiency and background rejections with this threshold are examined as a function of kinematic quantities, such as jet p_T or η . We also look at the “flat WP” where the b -tagging discriminant threshold is adjusted for each p_T (or η) bin, and then examine background rejection as a function of the jet p_T or η .

6.3 Low level taggers

To best identify the key physics features of b -jets, ATLAS b -taggers utilize two types of “low-level algorithms”: (1) impact parameter based (such as IP2D, IP3D and RNNIP) which use the individual track features (such as IPs) to classify the jet, and (2) vertex based algorithms (such as SV1 and JetFitter) which iteratively reconstruct the topology of the displayed decay. This section details these low-level algorithms, as shown in Figure 6.5. The outputs of these low-level algorithms are fed into another classifier, DL1r, which will be described after this section.

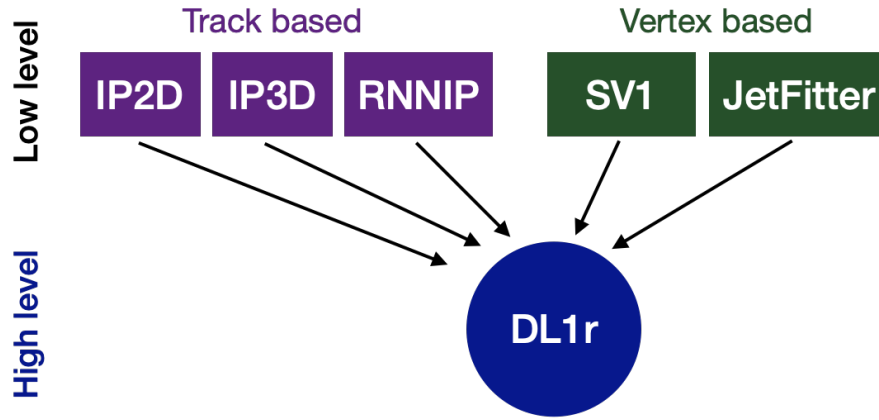


Figure 6.5: Types of b -taggers used on ATLAS

6.3.1 IP2D and IP3D

The IP2D and IP3D taggers look for tracks significantly displaced from the PV to identify b -jets. Crucial for identifying tracks from a long-lived decay is the “lifetime sign” assigned to the impact parameters of the track, where a positive lifetime indicates that the track is consistent with the hypothesis of being downstream of the flight path. The track is extrapolated back to the point of closest approach to the PV, and the displacement vector from the PV ($\Delta\vec{r}_{IP}$) is projected along the jet-axis. The sign of this projection vector is the “lifetime sign”, and the positive IP for a b -hadron decay is indicated by the red track in Figure 6.6. The tracks in light jets that originate from the PV are equally likely to be extrapolated above or behind the jet axis and are equally likely to have positive or negative lifetime sign. This positive lifetime sign from long-lived decays leads to an asymmetric IP distribution for the tracks in heavy flavor jets, shown in Figure 6.7.

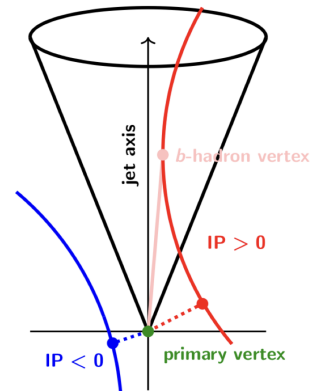


Figure 6.6: Lifetime signage illustration, positive for displaced decays [103].

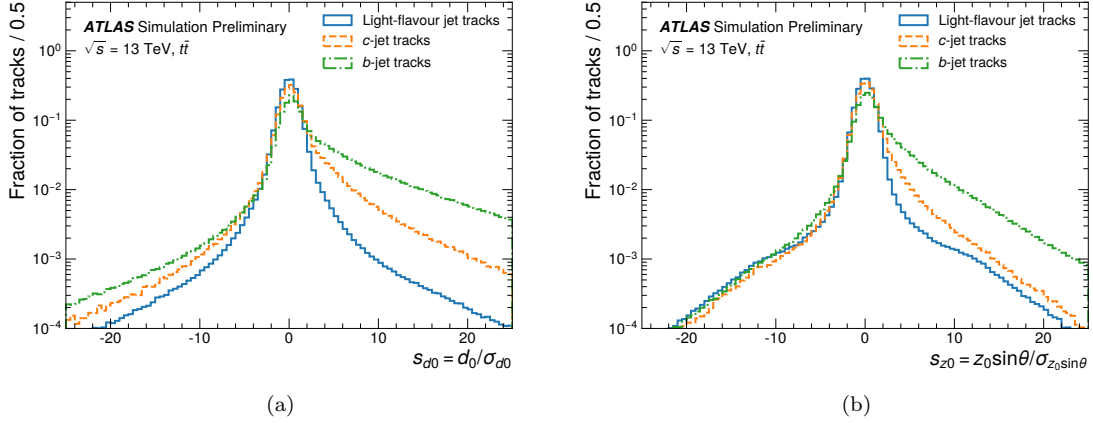


Figure 6.7: Lifetime signed transverse (a) and longitudinal (b) significances for b -jets, c -jets and light-flavor jets [5].

This lifetime sign can be calculated either in 3d, or with two 2d projections as given below [68].

3d sign

$$\text{sign}_{3D} = \text{sign} \left([\vec{p}_{trk} \times \vec{p}_{jet}] \cdot [\vec{p}_{trk} \times \Delta \vec{r}_{IP}] \right) \quad (6.3)$$

2d sign

$$\begin{aligned} \text{sign}_{r\phi} &= \text{sign} \left(\sin(\phi_{jet} - \phi_{trk}) \cdot d_{0,trk} \right) \\ \text{sign}_z &= \text{sign} \left((\eta_{jet} - \eta_{trk}) \cdot z_{0,trk} \right) \end{aligned} \quad (6.4)$$

Mathematical motivation

In a classification problem for a collection of n tracks, $X = \{X_1, \dots, X_n\}$, we want to learn the probability to have a given jet flavor, $Y = \{b, c, \text{light}\}$ through the **discriminative model** $p(Y|X_1, \dots, X_n)$. Sometimes it is more natural to learn the **generative model** of the track features given the jet class, with the likelihood $p(X_1, \dots, X_n|Y)$. These probabilities are proportional by Bayes rule:

$$p(Y|X_1, \dots, X_n) = \frac{p(X_1, \dots, X_n|Y)P(Y)}{p(X_1, \dots, X_n)} \sim P(X_1, \dots, X_n|Y). \quad (6.5)$$

The Naïve Bayes model makes a strong assumption that the tracks are independent of each other, given the jet class. This is expressed by Eq. 6.6. Independence is a strong restriction on the model's expressivity, but increases the predictive power by decreasing the number of parameters. Only the probability template of a single track $p(X_i|Y)$ needs to be learned.

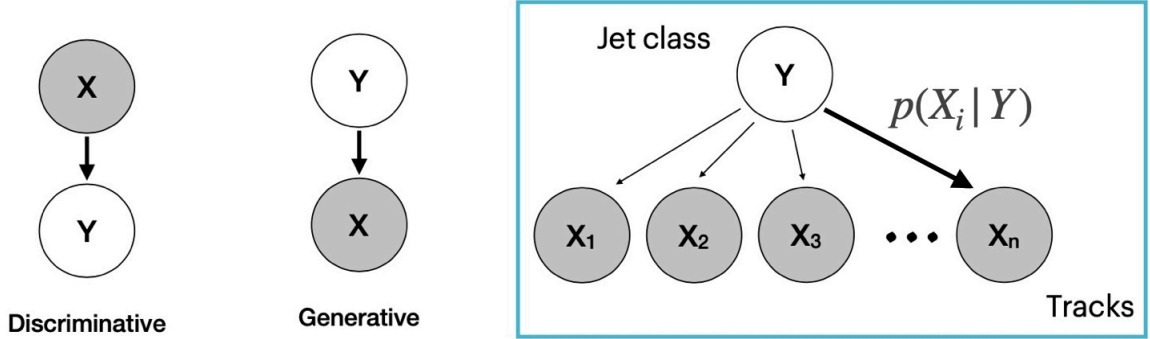


Figure 6.8: Different modeling assumptions for probabilistic graphical models. A Naïve Bayes model is a generative model that assumes that the tracks' features are *independent* [104].

$$p(Y|X_1, \dots, X_n) \sim P(X_1, \dots, X_n|Y) \approx \prod_{i=1}^n p(X_i|Y) \quad (6.6)$$

IPXD algorithms

The IP2D and IP3D algorithms are Naïve Bayes models which define the likelihoods $p(X_i|Y)$ based on the lifetime signed significances: $s_{d0} = d_0/\sigma_{d0}$ and $s_{z0} = z_0 \sin \theta / \sigma_{z_0 \sin \theta}$ [105]. The likelihoods also include 14 exclusive categories defined by the hit patterns of the tracks, shown in Table 6.2. IP3D builds three-dimensional likelihood templates from the tracks' s_{d0} and s_{z0} , and category, while IP2D builds the template based on the tracks' s_{d0} and category. On a technical note, these IP2D and IP3D significances use different definitions of the lifetime sign. IP2D uses the d_0 sign based on the 2d (x,y) projection, while IP3D uses d_0 signed with the 3D vectors, and z_0 signed based on the 2d projection. The inclusive distributions of the IP3D d_0 and $z_0 \sin \theta$ lifetime signed significances for the different jet flavors are shown in Figure 6.7. The IPXD uses a high purity track selection with $|d_0| < 1$ mm, $|z_0 \sin \theta| < 1.5$ mm, and $p_T > 1$ GeV.

The jet-level likelihoods can be constructed by multiplying the individual probabilities, the IP3D b -tagging discriminants are therefore defined as:

$$D_{\text{IP3D},l} = \log \prod_{i \in \text{tracks}} \frac{p_b^i}{p_l^i} \quad D_{\text{IP3D},c} = \log \prod_{i \in \text{tracks}} \frac{p_b^i}{p_c^i}. \quad (6.7)$$

The IP2D discriminants are defined analogously with the corresponding IP2D templates.

Figure 6.9 shows the correlation between the largest two $|s_{d0}|$ s tracks in b -jets (left) and light jets (right). Although these track significances are uncorrelated in light jets, there is a strong correlation in the track significances in b -jets since we expect five displaced tracks from the heavy flavor decay. The RNN-based IP algorithm (presented in the next section) is a compact way to represent the probability distribution $p(Y|X_1, \dots, X_n)$, and alleviate the need for the independence assumption

#	Category	Fractional contribution [%]		
		<i>b</i> -jets	<i>c</i> -jets	light-jets
0	No hits in first two layers; expected hit in IBL and b-layer	1.9	2.0	1.9
1	No hits in first two layers; expected hit in IBL and no expected hit in b-layer	0.1	0.1	0.1
2	No hits in first two layers; no expected hit in IBL and expected hit in b-layer	0.04	0.04	0.04
3	No hits in first two layers; no expected hit in IBL and b-layer	0.03	0.03	0.03
4	No hit in IBL; expected hit in IBL	2.4	2.3	2.1
5	No hit in IBL; no expected hit in IBL	1.0	1.0	0.9
6	No hit in b-layer; expected hit in b-layer	0.5	0.5	0.5
7	No hit in b-layer; no expected hit in b-layer	2.4	2.4	2.2
8	<i>Shared</i> hit in both IBL and b-layer	0.01	0.01	0.03
9	At least one <i>shared</i> pixel hits	2.0	1.7	1.5
10	Two or more <i>shared</i> SCT hits	3.2	3.0	2.7
11	<i>Split</i> hits in both IBL and b-layer	1.0	0.87	0.6
12	<i>Split</i> pixel hit	1.8	1.4	0.9
13	Good quality	83.6	84.8	86.4

Table 6.2: Categories for defining the IP2D and IP3D templates [106].

[93].

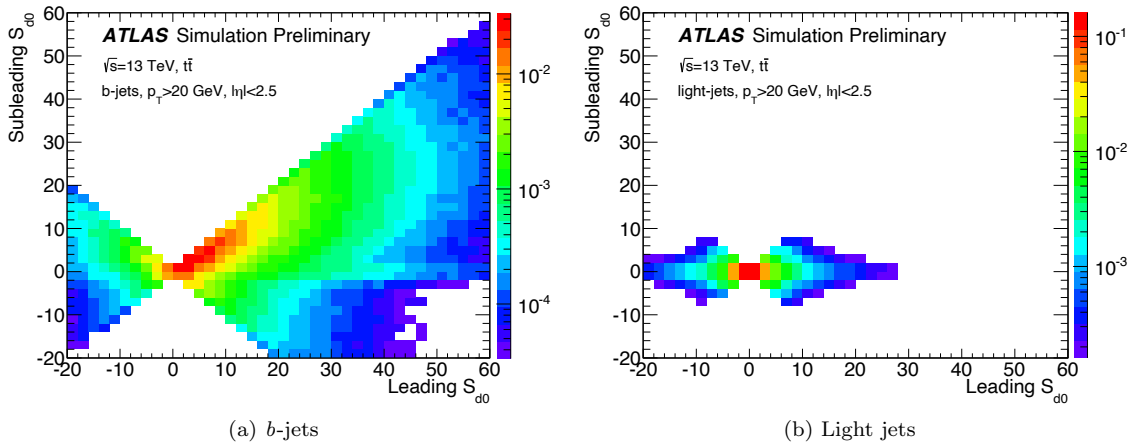


Figure 6.9: Correlation between the two leading significance tracks in b-jets (left) and light-jets (right) [93].

Additionally, IPXD scales poorly with adding more features for each track, since the number of bins for histograms defining the $p(X_i|Y)$ templates scales exponentially with the number of dimensions of X_i . Alleviating this curse of dimensionality problem is a natural introduction for ML-based solutions.

6.3.2 RNNIP

Recurrent Neural Networks (RNNs) operate on *sequences* by iterating over the elements and at each step, updating a fixed dimensional hidden state, as shown illustrated in Figure 6.10. The final fixed-dimensional vector can be used for classification. A common use case for RNNs is natural language processing, where the sequence elements are given by the order of the words in a sentence, so the sequence elements are often referred to as “time steps”. An RNN is a compact representation since the weights are shared between the time steps of the input processing. This also allows the RNN to operate natively over a variable number of inputs. A Long Short Term Memory (LSTM) cell for an RNN additionally adds a context to preserve long-range correlations between the elements of the sequence and avoid the problem of vanishing gradients for the update steps through the computational graph [107].

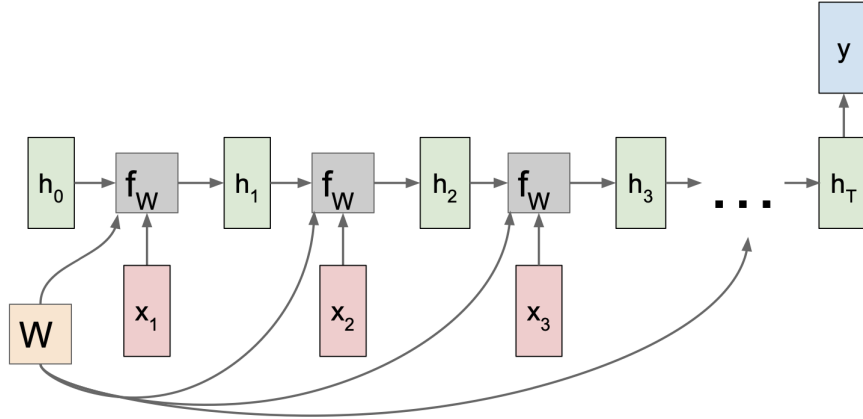


Figure 6.10: Illustration of how an RNN works. The $X_1, X_2, X_3 \dots$ are the inputs, the $h_i, i = 1, \dots, T$ is the hidden state of the RNN while the f_W is the function that updates the hidden state with the new input. The W are the weights which are common across the timesteps X_i . The sequence outputs a classification label y [108].

We use an RNN for b -tagging, by treating the tracks in the jet as a *sequence* where each track has a set of features X_i . The RNNIP is a b -tagging algorithm that uses an LSTM [109] to natively operate variable length tracks in a jet while taking into account the correlations between the tracks. As shown in [93], the accounting for these correlations allows the RNN to be more performant than IP3D even when trained on the same inputs. But with using NNs instead of histograms – we avoid the “curse of dimensionality” and can use additional variables sensitive to the kinematics of the b -hadron decay which significantly improves performance [93]. To represent the jet as a sequence, we needed to specify an ordering for the tracks, and the significance of the tracks, s_{d0} , was empirically found to be a good ordering for the track inputs. RNNIP uses the same tight track selection as IPXD with $|d_0| < 1$ mm, $|z_0 \sin \theta| < 1.5$ mm, and $p_T > 1$ GeV.

We’ve included further optimizations to the RNNIP algorithm with respect to [93]. We’ve

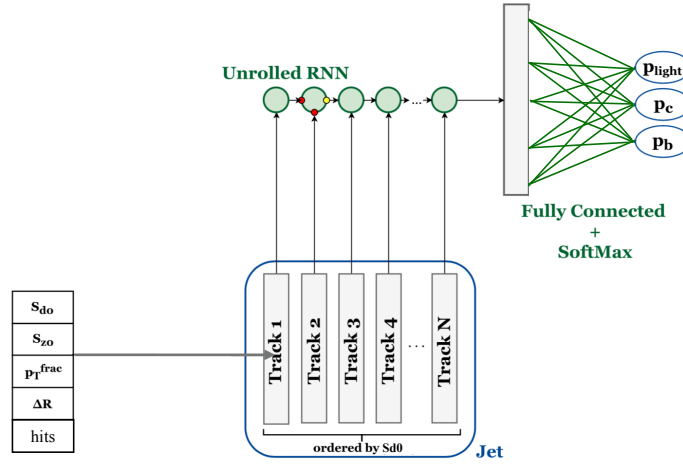


Figure 6.11: RNNIP architecture (modified from [93]).

increased the architecture size, and included a dropout layer [110] with a dropout fraction of 0.2 to improve the network’s generalization to the test set. Previously, the track category (Table 6.2) was used as an RNNIP input by encoding this category in a 2-dimensional embedding vector. The track category is defined based on hit patterns in the IBL and the next-to-innermost pixel layer (PIX1) due to their strong impact on the IP significance distributions. Now instead of using the category embedding, the “hits variables” which defined the IPXD categories are included as additional inputs. This additional low-level information led to an $\mathcal{O}(10\%)$ improvement in the l -jet background rejection. Table 6.3 summarizes the input features for the RNNIP tracks. Since the p_T^{frac} and ΔR variables have a tail at larger values, the natural log of the value for these variables is used as the feature to improve the time for the training to converge. Variable normalization to zero mean and unit variance is frequently used to preprocess features in ML algorithms. As many of our input variables already have near zero mean, only a subset of the track features are normalized: $\log p_T^{frac}$, $\log \Delta R$, nPixHits, nSCTHits.

To avoid classifying the jets just based on the differing kinematics of b , c , and light jets in $t\bar{t}$, the RNNIP training reweights the p_T spectra for b -jets and c -jets to match light jet p_T spectrum. These weights are then passed to the loss function in the RNNIP training. For the D_b definition for the RNNIP evaluations, $f_c = 0.07$ is used as the fraction of c -jets relative to non- b -jets in $t\bar{t}$ events.

RNNIP PFlow optimization

Figure 6.12 illustrates why b -tagging becomes harder at high p_T . As the jet p_T increases, there are more tracks from the fragmentation process. Additionally, as the b -hadron becomes more boosted, its decay products become more collimated making reconstructing the tracks in the dense environment at the core of the jet more challenging (as shown earlier in Figure 5.5). Finally, as the b -hadron

Input	Description
s_{d0}	d_0/σ_{d0} : Transverse IP significance
s_{z0}	$z_0 \sin \theta / \sigma_{z_0 \sin \theta}$: Longitudinal IP significance
$\log p_T^{frac}$	$\log p_T^{track} / p_T^{jet}$: Logarithm of fraction of the jet p_T carried by the track
$\log \Delta R$	Logarithm of opening angle between the track and the jet axis
IBL hits	Number of hits in the IBL: could be { 0, 1, or 2 }
PIX1 hits	Number of hits in the next-to-innermost pixel layer: could be { 0, 1, or 2 }
shared IBL hits	Number of shared hits in the IBL
split IBL hits	Number of split hits in the IBL
nPixHits	Combined number of hits in the pixel layers
shared pixel hits	Number of shared hits in the pixel layers
split pixel hits	Number of split hits in the pixel layers
nSCTHits	Combined number of hits in the SCT layers
shared SCT hits	Number of shared hits in the SCT layers

Table 6.3: Track features for the RNNIP algorithm.

becomes increasingly boosted, it can travel past the first layers of the detector before decaying, as shown in Figure 6.13. Missing the first hits on the track hurts the track reconstruction efficiency. The conflation of these factors is shown in Figure 6.12, where heavy flavor track reconstruction efficiency decreases as the jet p_T increases. The RNNIP track selection keeps less than one heavy flavor track (on average) for multi-TeV jets, making training maximally efficient taggers over a broad dynamic range a challenging problem.

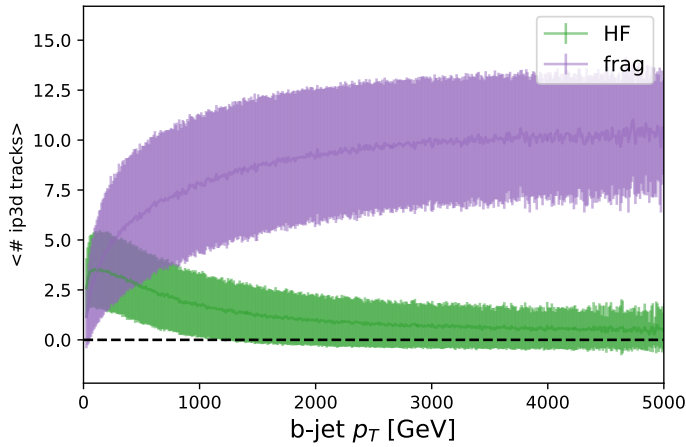


Figure 6.12: The evolution of heavy flavor compared to the number of fragmentation tracks that have $p_T > 1$ GeV, $|d_0| < 1$ mm, $|z_0 \sin \theta| < 1.5$ mm (the IP3D and RNNIP track selection).

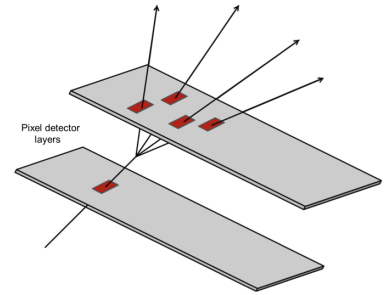


Figure 6.13: As the b -hadron becomes increasingly boosted, it will travel farther in the decay [111]

To produce an optimal tagger over this broad dynamic range, we trained RNNIP on five samples:

- $t\bar{t}$,
- 1.5 TeV hybrid sample ,
- 5 TeV hybrid sample,
- 1.5 TeV Z' sample, and
- 5 TeV Z' sample.

The model size was optimized to ensure the extended hybrid training performed as well as the dedicated $t\bar{t}$ and Z' trainings for both the $t\bar{t}$ and Z' sample evaluations. This optimal model had 400 hidden units in the LSTM cell, an increase from the 50 hidden units of [93]. The training was done with the adam optimizer [112] with a learning rate of 0.001. Five million jets were used with 20% held out as a validation set, and the training was stopped when the validation loss had not improved in the last 20 epochs.

Figure 6.14 shows the result of the training evaluating the performance on the Z' 5 TeV sample. The training on only $t\bar{t}$ has very poor performance when evaluated on this high p_T sample. The training with the hybrid 1.5 TeV sample gives better background rejection, while the hybrid 5 TeV sample gives the best performance.

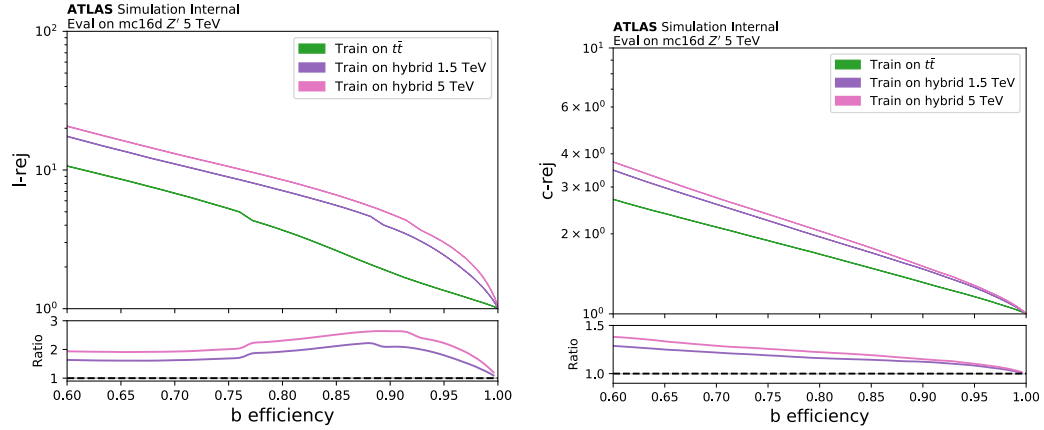


Figure 6.14: ROC curves for for light rejection (left) and charm rejection (right) evaluated on the 5 TeV Z' sample.

Figure 6.15 shows the “flat efficiency” background rejection as a function of the jet p_T . The dashed black boxes emphasize the high p_T region and demonstrate the robustness of the hybrid model to push the performance at high p_T . Both the hybrid 5 TeV and Z' 5 TeV samples are performing similarly for the highest p_T jets, as expected from an optimized tagger.

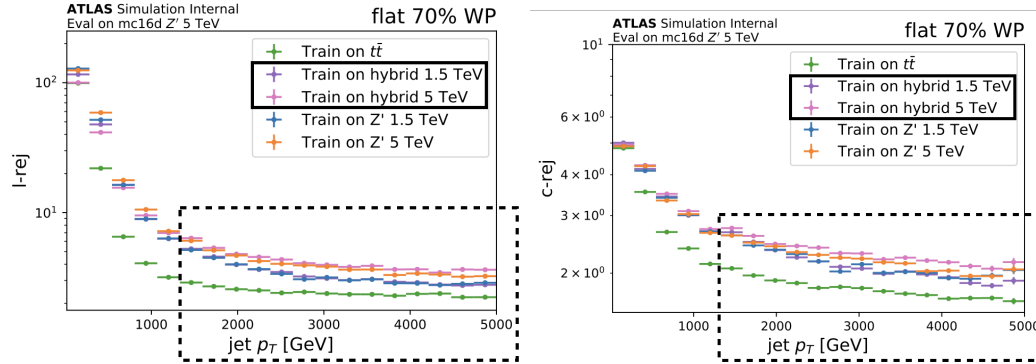


Figure 6.15: RNNIP performance as a function of p_T with the flat 70% WP for light rejection (left) and charm rejection (right). NN trained on 5 different samples are compared.

6.3.3 SV1

Given a set of tracks, a **vertex fit** calculates the vertex (x, y, z) location and fitted covariance matrix. This is done by minimizing the χ^2 of the residuals of the tracks extrapolated to the vertex location.

SV1 is a vertexing algorithm that reconstructs displaced decays with the assumption that all of the displaced tracks come from a single secondary vertex. Although this assumption is not strictly true as there's an underlying $b \rightarrow c$ cascade decay, in practice it helps make the problem tractable when the b and c hadron decays are close to each other, or when one of the weakly decaying b or c -hadrons only has a single track from the decay. Figure 6.16 shows jet displays which highlight three typical vertexing possibilities¹. The top panel shows a b -jet where the B and D hadron decays are each separately resolvable. The middle panel motivates the SV1 “single secondary vertex” assumption as the B and D hadron decays are not separately resolvable. The bottom panel shows a c -jet, where the c -hadron decay also results in a single secondary vertex, albeit less displaced from the primary vertex than was the case for the b -hadron decays in the b -jets.

The set of input tracks for SV1 have $|d_0| < 3.5\text{mm}$, $|z_0 \sin \theta| < 5\text{mm}$ and $p_T > 500\text{ MeV}$ s [68]. To reduce the rate of fake vertices, additional track cleaning cuts are applied for jets with $\eta > 1.5$, since these tracks pass through more material. To limit the fake rate in high p_T jets, only the 25 highest p_T tracks inside of the jet are considered [113]. These track pre-selection cuts are looser than the IPXD (and RNNIP) algorithms because the probability for these tracks to be attached to the fitted vertex provides an additional track selection criterion.

All possible pairs of 2-track vertices are formed which have a fitted χ^2 probability larger than 3.5%. These two track vertices require each track to have a 3d significance larger than 2σ and the sum of the 3d track significances to be larger than 6σ . Finally, the 2-track vertex needs to have a positive lifetime sign (i.e, a positive projection along the jet axis) [68]. Two track vertices consistent

¹Many thanks to Jonathan Shlomi for the nice track extrapolator code.

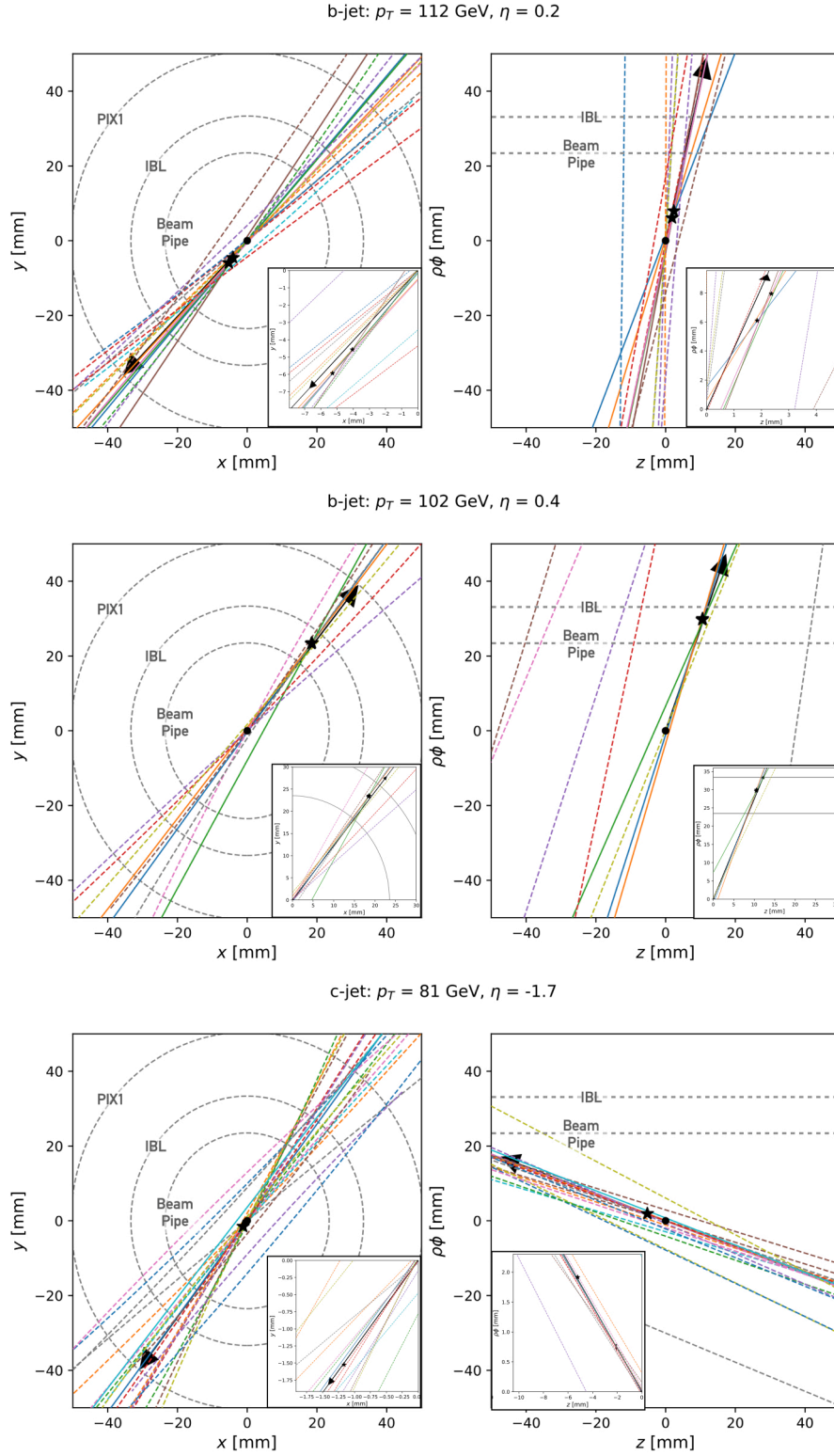


Figure 6.16: Visualization of the (x, y) and $(z, \rho\phi)$ 2d views of the tracks for reconstructing the secondary vertex (or vertices). The arrow on the figure indicates the jet axis, and a \star shows where each weakly decaying hadrons. The solid lines are tracks from the heavy flavor decays, while the dashed lines show the other tracks associated to the jet.

with K_s and λ^0 decays and e^+e^- conversions are vetoed by looking at the $m_{\pi^+\pi^-}$, $m_{p\pi^-}$, and m_{ee} distributions, respectively.² Two track vertices from hadronic material interactions are vetoed if vertex position overlaps with the detector material.

This set of two track vertices gives a “cleaned” set of tracks that are then used to fit a single SV. The vertex fit is done, and the mass of the SV is calculated using a π^\pm mass for the tracks then using the four-vector sum of the tracks to calculate an invariant mass. If the χ^2 probability of the fit is less than 0.1%, or the SV mass is larger than 6 GeV, the track with the largest track-vertex association χ^2 is removed. This vertex fit is repeated iteratively, removing tracks until the convergence criterion is satisfied.

For a true displaced decay, this secondary vertex will have properties consistent with a B or D hadron, and some key discriminating variables include the mass of the secondary vertex, the energy fraction of the vertex relative to the jet, the displacement of the fitted SV from the PV, and the track multiplicity of the fitted vertex. The list of the eight SV1 inputs that are fed into the high level b -tagger is enumerated in Table 6.4, and a subset of these features are shown in Figure 6.17.

Input	Description
m	Invariant mass of the tracks reconstructed in the secondary vertex
f_E	The energy of the SV over the energy of the jet
$\Delta R(\vec{p}_{jet}, \vec{r}_{SV} - \vec{r}_{PV})$	Opening angle between the jet and the SV flight axis
L_{xy}	SV transverse distance from the PV
L_{xyz}	SV distance from the PV
S_{xyz}	Significance of the displacement of the SV: L_{xyz}/σ_{xyz}
N_{trk}	Number of tracks in the SV
$N_{2-trk-vtx}$	Number of 2 track vertices before the vertex fit

Table 6.4: Features from the SV1 reconstruction that are fed as input to the DL1r tagger.

²The tracker only measures the momentum of the particle, so to reconstruct a mass, we need to make a particle ID assumption for the track’s mass to get the 4-vector. The meson (or lepton) mass used for each track to reconstruct the 2-track vertex invariant mass is denoted by the subscripts. For the $\Lambda^0 \rightarrow p\pi^-$ reconstruction, the leading p_T track is taken as the p track.

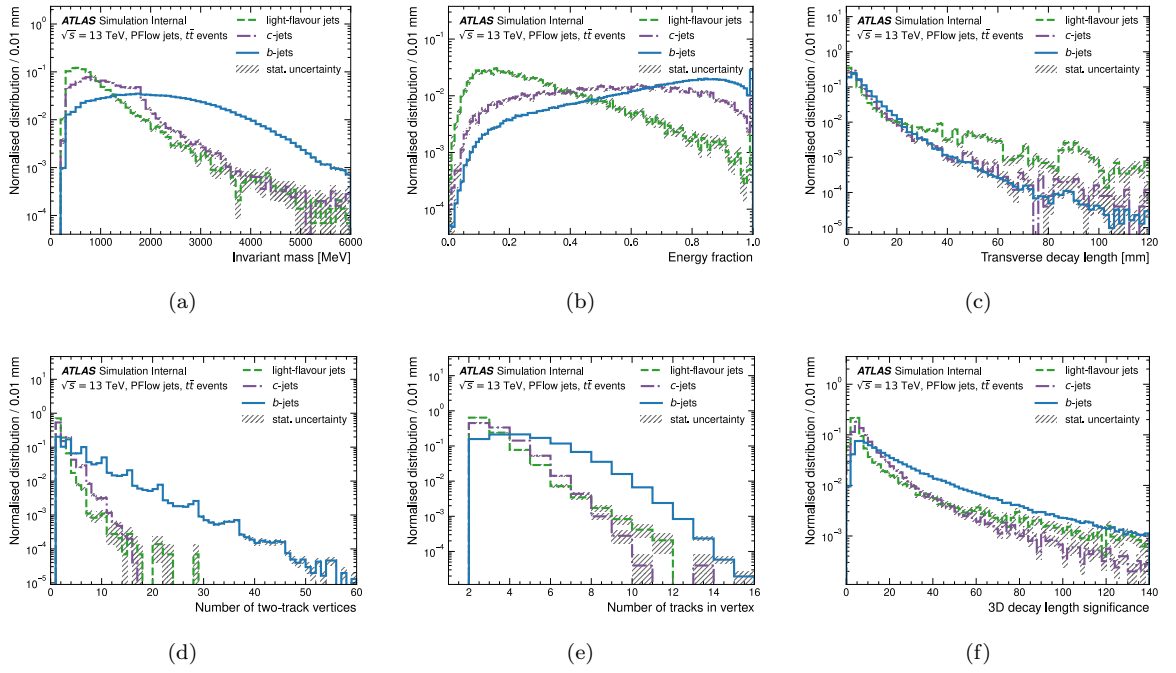


Figure 6.17: A selection of the SV1 features that are inputs for the high-level DL1r *b*-tagger [3].

6.3.4 JetFitter

JetFitter (JF) is the other ATLAS vertexing algorithm that fits multiple secondary vertices to better represent the underlying cascade topology visualized in Figure 6.18. The key assumption this algorithm makes to limit the complexity of the multi-vertex fit is that all of the displaced vertices lie on the same flight axis. This is well motivated as the B -hadron carries a large portion of the initial quarks momentum which causes the B (and subsequent D) hadron to have their momentum aligned with the original b -quark.

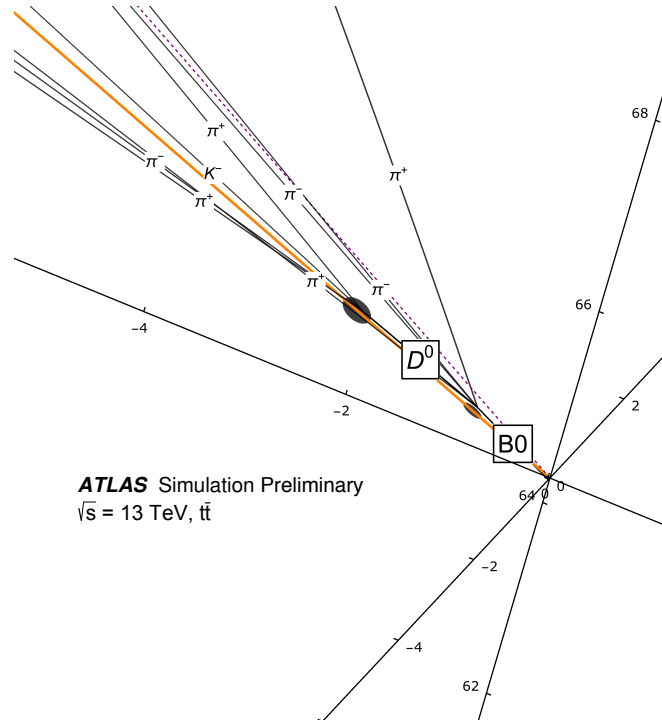


Figure 6.18: Visualization of the cascade topology of $B^0 \rightarrow D^0$ decay. The axes on the plot are in mm. The hadron decays are shown with grey circles. The common B^0 / D^0 flight axis is shown in the orange line, which is initialized with the jet axis, shown by the purple dashed line. [114]

The tracks input to the JF algorithm have a loose selection criteria of $|d_0| < 3.5$ mm and $|z_0 \sin \theta| < 1.5$ mm, and $p_T > 500$ MeV [68]. Tracks compatible with originating from the PV are vetoed, with a harder cut for the tracks with a negative lifetime sign. The track's χ^2_i contribution to the PV fit is used to calculate the χ^2 probability for the PV compatibility, $\text{Prob}(\chi^2)$. For positively signed tracks, $\text{Prob}(\chi^2) < 10\%$, and $\text{Prob}(\chi^2) < 5\%$ for negatively signed tracks [114]. All 2-track vertices are formed with a χ^2 fit probability $> 3\%$ (5%) and a 3d signed significance $L_{3D}/\sigma_{3D} > 1.5(1)$.³ Tracks consistent with hadronic interactions are vetoed by removing tracks in

³The parentheses indicate the tighter vertex fit probability cut for smaller 3d significances.

2-track vertices that overlap with the detector material.

The multi-vertex fit extends the Kalman Filter formalism with the extended state vector of the Hidden Markov Model given by Eq. 6.8. The common flight axis is characterized by the (ϕ, θ) , and is initialized with the jet axis. Then the N displaced vertices are identified by the distance along this flight path from the PV (denoted by the d_1, d_2, \dots, d_N distances).

$$\vec{d} = (x_{PV}, y_{PV}, z_{PV}, \phi, \theta, d_1, d_2, \dots, d_N) \quad (6.8)$$

Initially, N is the number of input tracks. After the first fit for the $(\phi, \theta, d_1, d_2, \dots, d_N)$, the vertices with a χ^2 vertex fit probability less than 0.1% are removed. Then the probabilities for any two vertices to be merged are computed by redoing the vertex fit merging each pair of vertices and computing χ^2 merged vertex fit probability. This defines a “table of probabilities” for all two vertex combinations. An iterative clustering procedure merges the vertices with the highest probability, refits the state vector, and then updates the table of probabilities with the updated vertex list. This vertex clustering continues until all the merging probabilities are above a threshold.

The JetFitter algorithm additionally has the flexibility to reconstruct the B and D -hadron decays from single track vertices with a second “track-reattachment” step. This step considers tracks not in the two-track vertices, but which satisfy a stricter selection criteria of $|d_0| < 1.5$ mm and $|z_0 \sin \theta| < 3$ mm, and $p_T > 750$ MeV. To reduce pile-up tracks, tracks with $|z_0 \sin \theta| / \sigma_{z_0 \sin \theta} > 5$ and $d_0 / \sigma_{d_0} < 2$ are additionally vetoed. The $(\phi, \theta, d_1, d_2, \dots, d_N)$ from the above fit is then extended with these extra single track vertices, and the vertex clustering procedure is repeated until convergence.

The set of input variables that characterize the result of the multi-vertex fit are listed in Table 6.5. There are two types of input variables. The first 8 variables listed in Table 6.5 characterize the properties of the tracks in the full cascade fit. The mass of the tracks associated to the displaced vertices, m_{JF} , includes a conservation of momentum constraint to account for the neutral tracks and improve the mass resolution (see Appendix A.2 for further details). A subset of these JF variables is shown in Figure 6.19.

The second block of variables enhance the sensitivity for c -tagging. For c -jets, only one displaced vertex is expected so these variables are only calculated when JF finds *exactly one* displaced vertex. These variables quantify the (2d and 3d) distance of the SV, its invariant mass, the energy of the tracks and track multiplicity of the displaced vertex. Additionally, since the c -hadron has a lower decay multiplicity than the b -hadron decay, each of the c -hadron tracks carry a larger proportion of the energy and hence have a wider opening angle [106]. The min, max, and avg η of the tracks in the jet and the tracks in the displaced vertex and the tracks in the jet characterize this wider opening angle.

Input	Description
m_{JF}	Invariant mass of the tracks attached to displaced vertices
f_E	Fraction of the energy in the displaced vertices compared to the jet's energy
$\Delta R(\vec{p}_{jet}, \vec{p}_{vtx})$	
S_{xyz}	Average significance of all of the displaced vertices
N_{trk}	Number of tracks associated to the fitted displaced vertices along the cascade decay chain
$N_{2-trk-vtx}$	Number of 2 track vertices (before the decay chain fit)
$N_{1-trk-vtx}$	Number of the single track vertices (after the decay chain fit)
$N_{\geq 2-trk-vtx}$	Number of the multi-prong displaced vertices (after the decay chain fit)
$L_{xyz}(SV)$	3d distance from the first displaced vertex
$L_{xy}(SV)$	transverse distance from the first displaced vertex
$m_{trk}(SV)$	Mass of the tracks associated to the first displaced vertex
$E_{trk}(SV)$	Energy of the tracks in the first displaced vertex
$f_E(SV)$	Energy fraction of the tracks in the first displaced vertex compared to the energy of the jet
$N_{trk}(SV)$	Number of tracks attached to the first displaced vertex
$\eta_{trk}^{max, min, avg}(SV)$	The maximum, minimum, and average pseudorapidity of the tracks in the displaced vertex
$\eta_{trk}^{max, min, avg}$	The maximum, minimum, and average pseudorapidity of the tracks in the jet

Table 6.5: Features from the JF reconstruction that are fed as input to the DL1r tagger. The first block of variables quantifies the global properties of the displaced vertices and decay topology. The second set of variables just looks at the properties of the first displaced vertex which capture the differences between b -jets and c -jets.

6.3.5 Low-level tagger comparison

Figure 6.20(a) compares the performance of the newly retrained RNNIP against some other low-level taggers. The SVKine (JFKine) line shows a small NN retrained with the jet p_T , η , and the SV1 (JF) inputs. RNNIP is the most efficient low-level algorithm at the highest efficiencies, although at the tightest efficiency 60% WP, SVKine has slightly better at light rejection and JFKine is slightly better at charm rejection. The blue line shows the DL1r performance and illustrates the gain from aggregating the complementary information of the different low level algorithms' assumptions. Figure 6.20(b) shows the corresponding evaluation on the Z' 5 TeV sample, demonstrating that RNNIP drives the performance at high momentum as the vertexing algorithms suffer when the HF track reconstruction efficiency is so low and the fragmentation tracks proliferate.

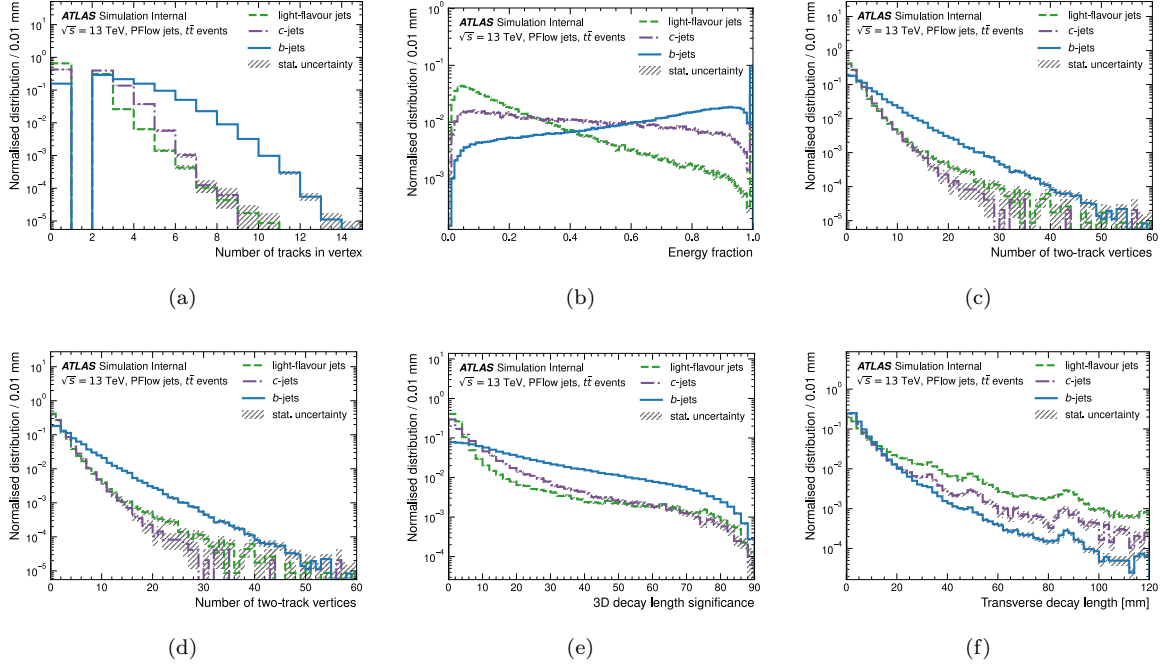
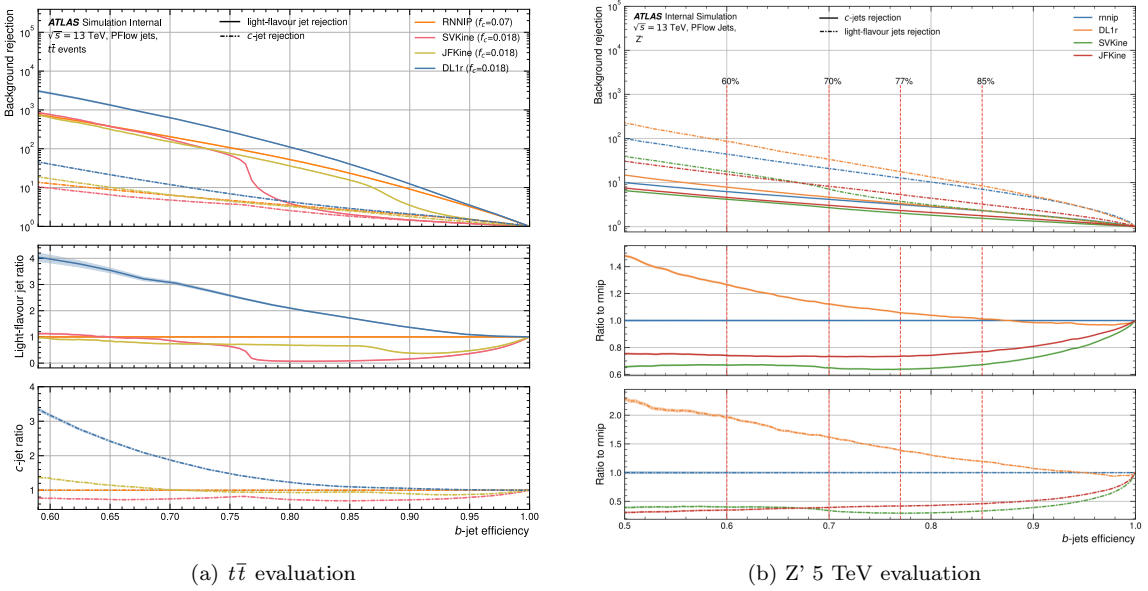
Figure 6.19: A selection of the JF features that are inputs for the high-level DL1r b -tagger [3].

Figure 6.20: Comparison of the performance of the low-level taggers [3].

6.4 DL1r High Level tagger

The Run 2 ATLAS b -tagger is the DL1r algorithm⁴. It is a feed-forward NN which includes inputs from the low-level taggers described previously, with the DL1r inputs summarized below:

- The jet p_T and η : to use the correlations between the jet’s kinematics and the low-level taggers
- **IP2D** and **IP3D**: The $\log p_b/p_c$, $\log p_b/p_l$, and $\log p_c/p_l$ are the DL1r NN inputs (2 IP taggers x 3 inputs = 6 inputs)
- **RNNIP**: After optimizing the RNNP hyperparameters, the DL1r algorithm accepted as inputs the RNNIP outputs: p_b , p_c , and p_l (3 inputs)
- **SV1**: vertex information (variables in Table 6.4) (8 inputs)
- **JF**: displaced vertices characteristics, Table 6.5 (20 inputs).

The low-level algorithms don’t always “succeed” or have a non-trivial output (i.e, for some jets, no tracks are selected by the tight IPXD selection cuts, or the vertex fit could fail to find a secondary vertex). When a low-level algorithm has a trivial output, default values are defined by setting the variable to the mean of the distribution (to be uninformative for the NN discrimination). The default value is set to zero if that is more physically meaningful, such as the mass or the energy of the SV is zero if no SV is found. To let the DL1r NN know if default values are used, boolean variables inputs are added for each low-level algorithm: **IP2D_isDefault**, **IP3D_isDefault**, **SV1_isDefault**, **JF_isDefault**, and **JFc_isDefault** (since the c -tagging variables are only calculated when JF finds exactly one SV. The DL1r tagger thus has 44 inputs:

$$\# \text{ DL1r inputs} = 2 (\text{jet } p_T, \eta) + 6 (\text{IPXD}) + 3 (\text{RNNIP}) + 8 (\text{SV1}) + 20 (\text{JF}) + 5 (\text{boolean defaults}) = 44.$$

These inputs are then pre-processed to have zero mean and unit variance (except for the boolean default indication variables).

The architecture has eight hidden layers with 256, 128, 60, 48, 36, 24, 12, and 6 hidden nodes per layer, and ReLU non-linearities.⁵ A softmax output layer has three output nodes for the b , c , and light jets. The categorical cross-entropy loss is used, and the parameters of the NN are optimized with a stochastic gradient descent optimizer adam [112] with a learning rate of 0.01, and a batch size of 15,000 jets.

The DL1r training ameliorates the jet class kinematic dependence by resampling the input jets. The input training sample is discretized into bins of jet p_T and η , and different jet flavors. Samples are drawn from this histogram to give a uniform number of jets across the (p_T, η) spectrum. 22 million jets are used for the input training for DL1r, with an equal proportion of b , c , and light jets. Then f_c was optimized to increase the charm rejection, while avoiding any degradation in the light jet

⁴The “r” in the DL1r algorithm name refers to the inclusion of the RNNIP information

⁵The performance was stable for a range of choices in the number of nodes and number of layers.

rejection relative to the EMTopo DL1 training. The DL1r performance plots thus use $f_c = 0.018$ in the D_b definition.

The comparisons of the DL1r tagger are made against two other taggers: (1) DL1, which uses the same inputs as DL1r except for the RNNIP variables, and (2) MV2, which uses the same inputs as DL1, but is a BDT instead of a NN. The DL1r tagger also has a reduced memory footprint compared to the previous MV2 tagger [3].

6.4.1 PFlow optimization

Figure 6.22 shows the inclusive ROC curves of the high-level taggers evaluated on $t\bar{t}$. All three of these taggers have been retrained on the PFlow jets, but the DL1r curve includes the RNNIP outputs from the extended hybrid training. The gains from RNNIP are up to a factor of 2 increase in the light jet rejection and up to 50% better at charm jet rejection. In Figure 6.21(a) the stars emphasize that the light jet rejection at the 70% WP for MV2 is similar to the charm rejection for DL1r at the 77% WP. All of the latest HH analyses use this new DL1r algorithm, and were able to take advantage of the better signal efficiency with this looser WP, translating into a direct improvement in the physics results.

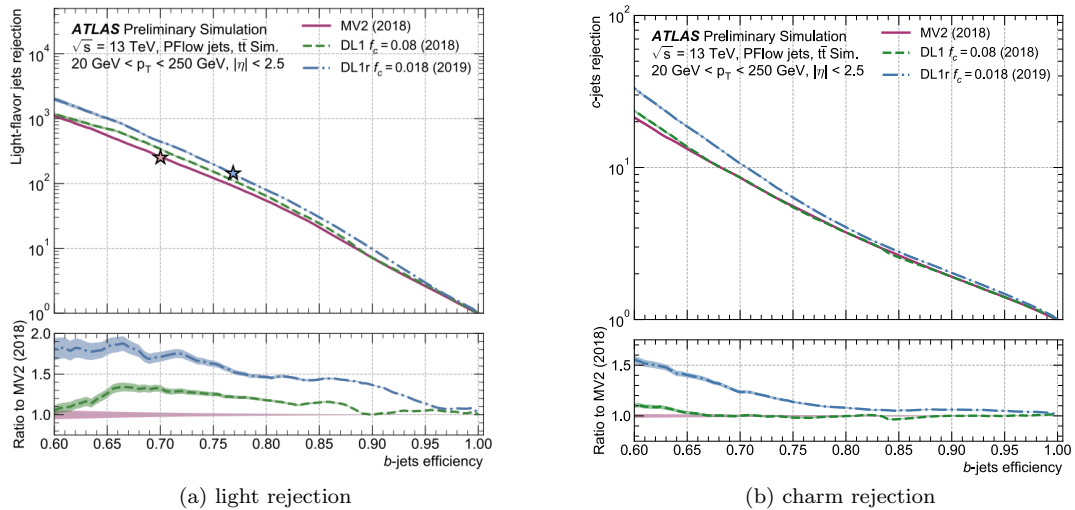


Figure 6.21: The ROC curves for light and charm rejection evaluated in $t\bar{t}$ comparing the PFlow retrained high-level taggers: MV2 (BDT-based), DL1 (NN based, RNNIP not included), and DL1r (Run 2 recommendation) [115]. The stars indicate the 77% WP for DL1r and the 70% WP for MV2.

Figure 6.22 shows the performance at the fixed 77% WP as a function of the jet p_T . In the $t\bar{t}$ p_T range, the b -tagging efficiency increases with the jet p_T . The b tagging efficiency is defined for each of the MV2, DL1, and DL1r taggers, giving these taggers similar b -tagging efficiency profiles. The gains of the DL1r tagger are seen with the background rejection. The RNNIP algorithm contributes

the most at high p_T as the improvement of DL1r is most evident for jets with p_T above 50 GeV.

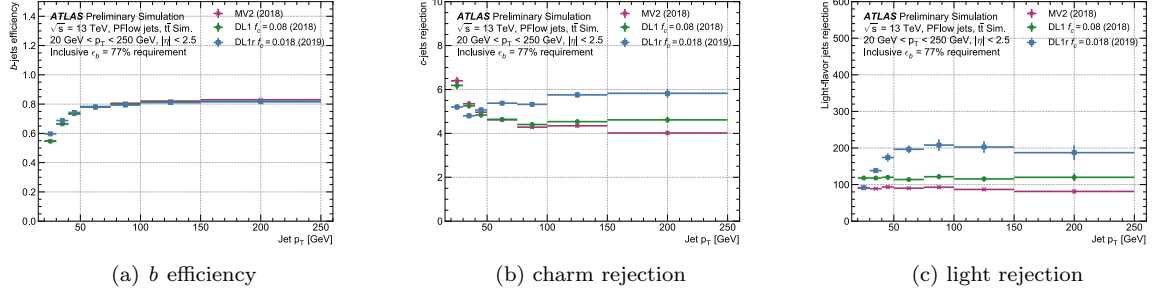


Figure 6.22: The p_T dependence of the b -tagging efficiency and background rejection in $t\bar{t}$ for the fixed 77% WP for the PFlow retrained high-level taggers [115].

For the flat 77% WP, the DL1r shows an uncontroversial outperformance of the DL1 and MV2 taggers. Although the “flat efficiency” plots are useful to understand the performance of the tagger, these flat working points are not calibrated.

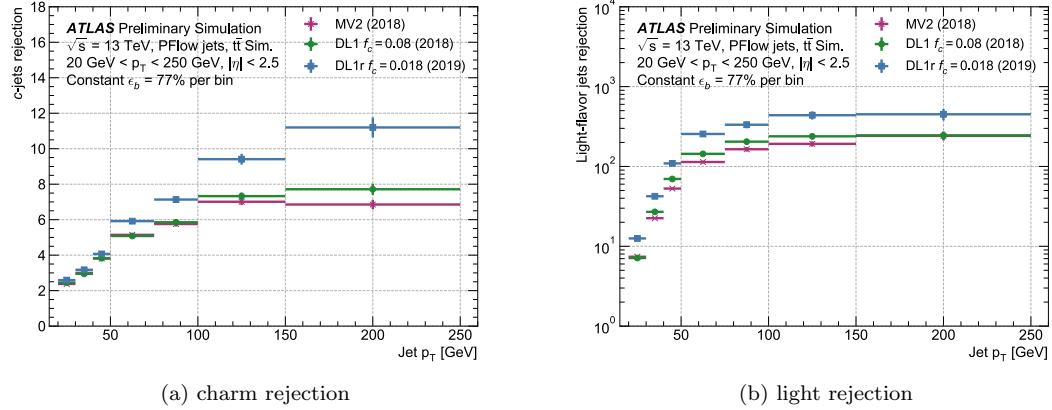


Figure 6.23: Background rejection in $t\bar{t}$ for the flat 77% WP for the PFlow retrained high-level taggers [115].

Figure 6.24 shows the DL1r performance in the 5 TeV Z' sample to examine the tagger performance extended to higher p_T . The blue curve shows the new DL1r training, while the orange compares to an older training. The newly retrained tagger has improved the b -jet efficiency up to 5 TeV jets (Figure 6.24(a)). The light jet rejection decreases as the jet p_T increases, as expected from the increase in the fragmentation tracks. Above 5 TeV, the b -jet rejection degrades, but helps the b vs light jet discrimination power as both the b -efficiency and light rejection improve.

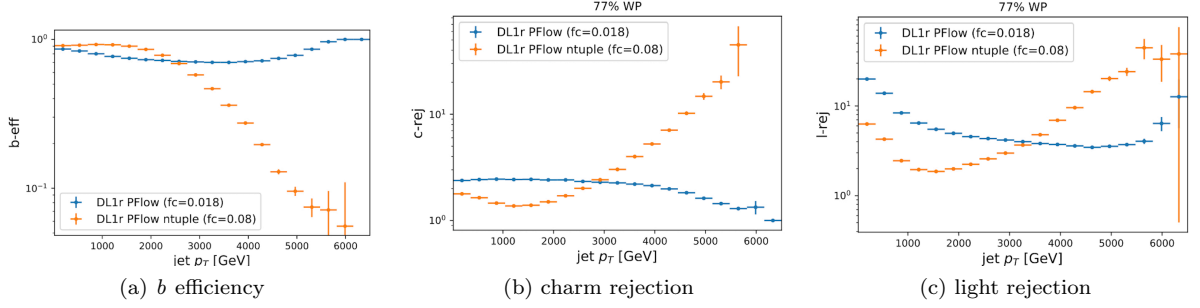


Figure 6.24: The p_T dependence in the 5 TeV Z' sample of the b -tagging efficiency and background rejection for the fixed 77% WP for the PFlow retrained high-level taggers (blue) compared to the old DL1r tagger in orange.

6.4.2 VR track jets optimization

Here we show the same performance plots for the high-level taggers. The comparison shows all the high-level taggers retrained for the VR track jet collection – which means all of these taggers are novel as the first time these taggers had dedicated b -tagging optimizations for this track jet collection. The comparison is still interesting to highlight the gains specifically from RNNIP. Figure 6.25 shows the ROC curves comparing these VR track jet trainings. The DL1r tagger again offers a factor of 2 improvement in the light rejection and up to 50% improvement for the charm rejection compared to the BDT-based MV2 tagger in a $t\bar{t}$ sample.

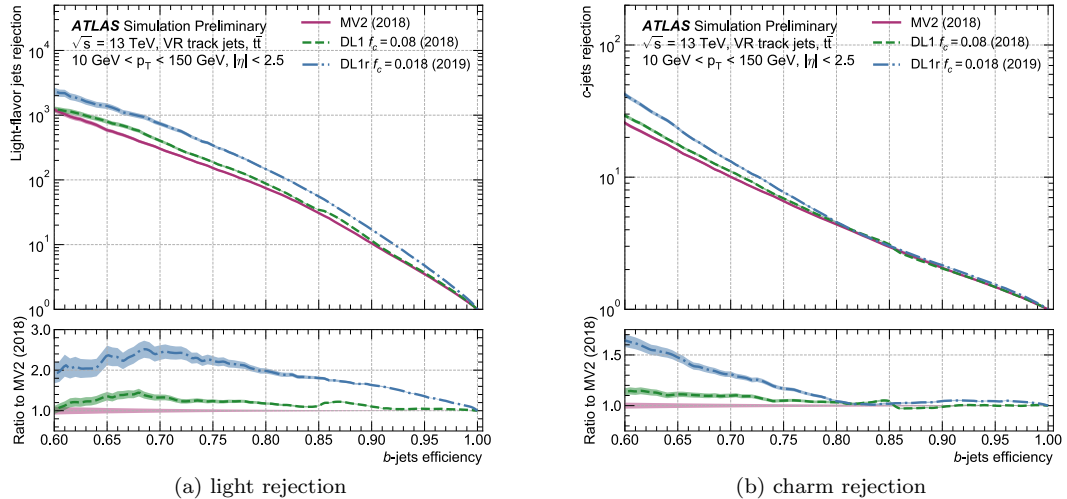


Figure 6.25: The ROC curves light and charm rejection comparing the VR retrained high-level taggers: MV2 (BDT-based), DL1 (NN based, RNNIP not included), and DL1r (Run 2 recommendation) [1].

The performance plots versus p_T for the fixed and flat WPs are shown in Figure 6.26 and Figure 6.27, respectively. We again see a more dramatic improvement offered by the RNN at the higher jet p_T s.

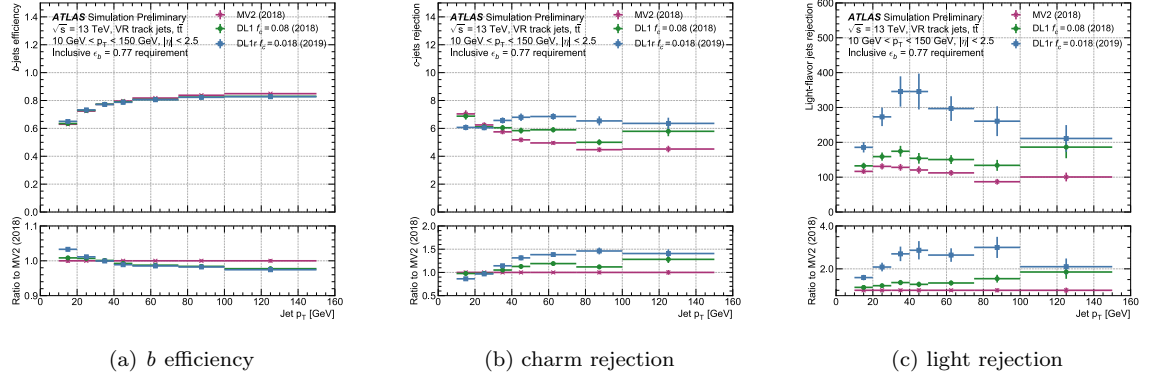


Figure 6.26: The p_T dependence of the b -tagging efficiency and background rejection for the fixed 77% WP for the VR retrained high-level taggers [1].

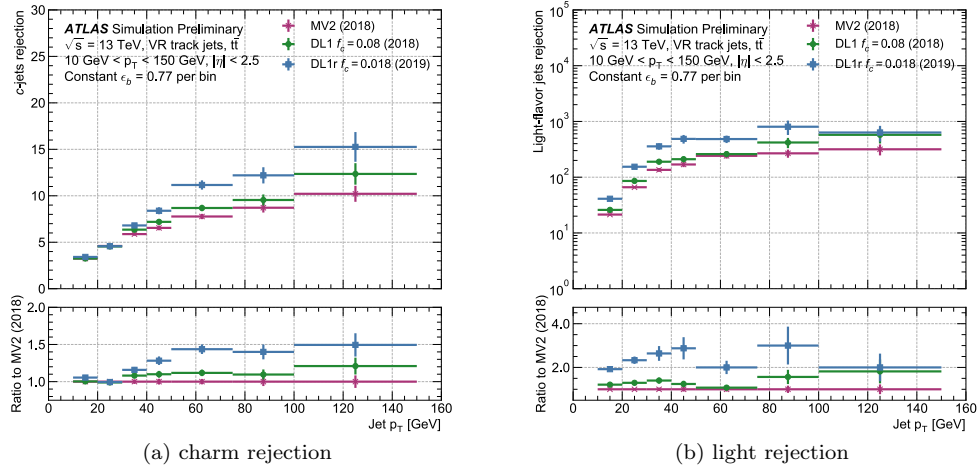


Figure 6.27: Background rejection for the flat 77% WP for the PFlow retrained high-level taggers [1].

Impact of FTAG improvements on analyses

Finally, the physics improvements from this VR track jet retraining is highlighted by the $HH \rightarrow 4b$ boosted analysis search in Figure 6.28. The 36 fb^{-1} result (shown in the dashed gray line) used fixed $R=0.2$ jets with the (EMTopo trained) MV2 algorithm at the 70% WP [116]. The recent result from the 139 fb^{-1} full Run 2 data set is shown in the blue dashed line which improves on the old result by using variable radius track jets, and a new dedicated track jet b -tagger training for the DL1r tagger. The signal efficiency was improved by using the looser 77% WP. This result is compared to what is expected just from the dataset scaling shown in the dashed black line. There is a nice improvement over our luminosity scaling expectations for scalar masses above 1250 GeV (where the boosted analysis drives the sensitivity). This improvement is larger as the heavy scalar mass increases, emphasizing the gains from the shrinking radii as the b -quarks from the Higgs decay become more collimated.

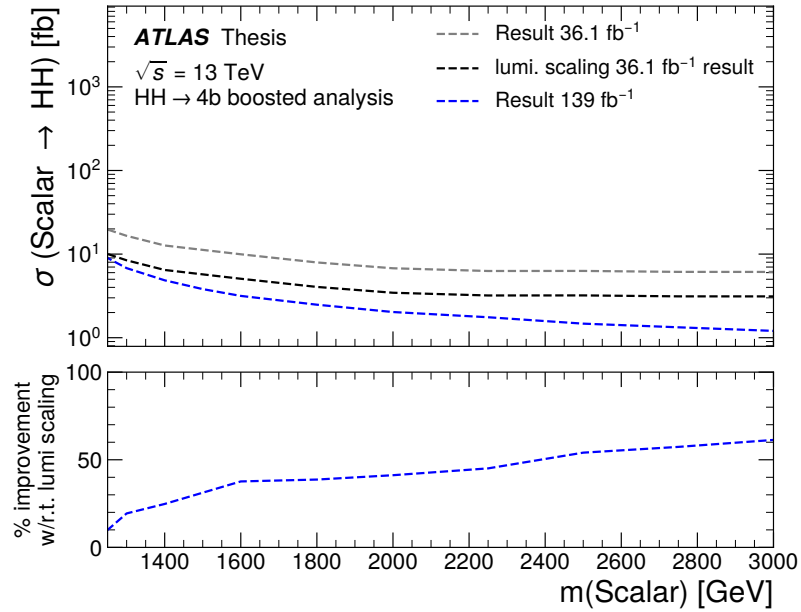


Figure 6.28: Demonstration of The grey line shows the 36.1 fb^{-1} result with fixed radius $R = 0.2$ track jets from [116]. The black line shows the expected improvement of this result expected just from the larger size of the 139 fb^{-1} dataset. The blue line shows the 139 fb^{-1} result from [117] which shows gains over the expectation from the *dedicated b -tagger retraining* for VR track jets.

6.5 FTAG Calibrations

While simulation gives the truth labels for the jet to train and evaluate the algorithm's performance, efficiency must ultimately be calibrated to data. Control samples are built with selections targeting a jet flavor, and the observed and simulated efficiencies are compared. The correction factors are called **scale factors** (SF) and defined as:

$$SF_f(p_T) = \frac{\varepsilon_{f,\text{data}}(p_T)}{\varepsilon_{f,\text{MC}}(p_T)}, \quad (6.9)$$

where $f = b, c$, or light, as these SFs are derived separately for each jet flavor, and these calibrations are summarized below.

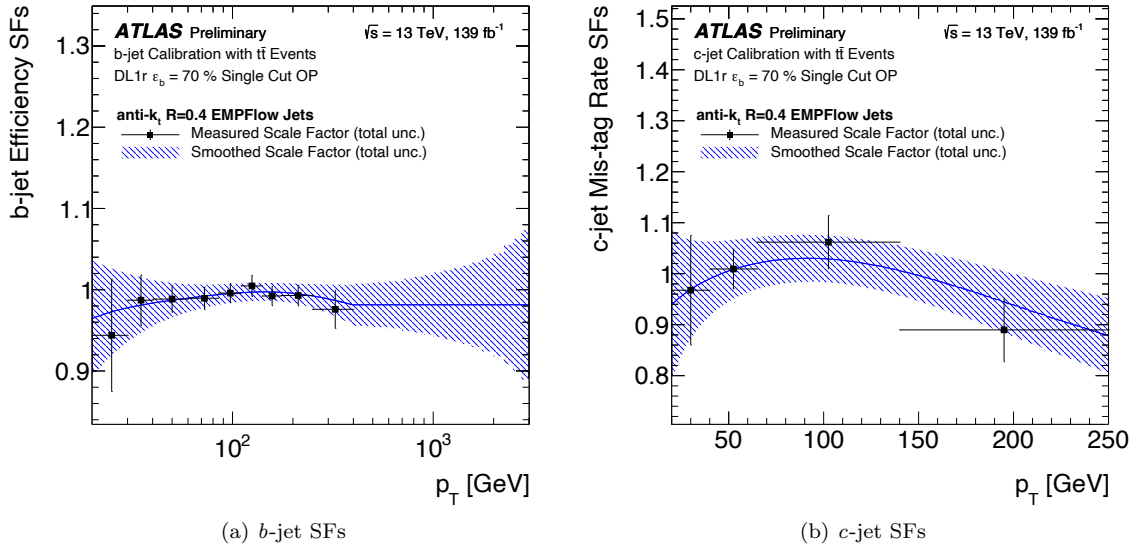


Figure 6.29: SFs for the b and c -jets for the DL1r tagger at the 77% operating point [2].

b -jet calibration

A dedicated $t\bar{t}$ selection is used to get a high purity sample of b -jets, illustrated in Figure 6.30. Exactly two jets are required, and di-leptonic decays are targeted by requiring exactly two leptons with opposite charge and opposite flavor. The invariant masses of the lepton, jet pairs, $m_{j1,\ell}$ and $m_{j2,\ell}$, are used to define CRs to constrain the flavor fractions in the fit [105]. These SFs are measured in nine p_T bins and the result is shown in Figure 6.29(a).

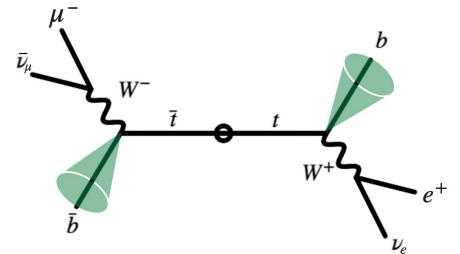


Figure 6.30: $t\bar{t}$ signature for the b -jet calibration. The green cones indicate the calibration b -jets.

When $t\bar{t}$ sample runs out of statistics for jets with $p_T > 400\text{GeV}$, the central value of the last p_T bin of the calibration is used for higher p_T jets, and the error bar is inflated with an MC-based extrapolation to account for (1) the decreased track reconstruction efficiency in the denser environments, (2) the jet energy scale and resolution, and (3) the b -hadron interaction model in the detector [118].

c -jet calibration

A dedicated semi-leptonic $t\bar{t}$ selection selects a sample of c -jets coming from the hadronic W decay, as shown in Figure 6.31. Exactly one muon or lepton is required with an $E_T^{miss} > 20$, and a kinematic likelihood method matches the jets to the partons [119]. The c -jet SFs are extracted in four p_T bins, and the result is shown in Figure 6.29(b).

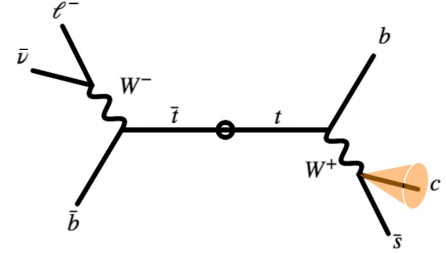


Figure 6.31: $t\bar{t}$ signature for the c -jet calibration. The orange cone indicates the calibration c -jet.

light-jet calibration

To select a pure sample of light-jets, a Z +jets sample is used where the Z is identified from decays to electrons or muons [4]. Figure 6.32 gives a representative diagram for light jets produced in association with the Z . However, it is difficult to get a pure sample of light-jets after applying a b -tagging cut because of how powerful the b -taggers are at rejecting light jet backgrounds. Figure 6.21(a) shows that for at the (tightest) 60% WP, only 1 out of 2000 light-jets will be mis-tagged. This makes it impossible to measure light-jet yield in data directly as the sample becomes dominated by heavy flavor jets after applying a b -tagging cut.

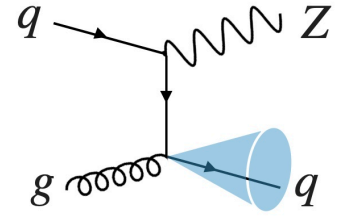


Figure 6.32: Z +jets representative diagram for the light jet calibration. The blue cone indicates the calibration light jet.

Instead of calibrating the light mistag rate directly, a “flipped” tagger is calibrated which has:

1. the same performance for light jets, but
2. reduced efficiency for tagging b and c -jets.

A large fraction of light-flavour jets are wrongly classified as b -jets when there is a track with a large significance. This could come from the track being mismeasured because of the detector resolution, or because the tracks originated from a pile-up vertex. Both these effects have an equal probability to give a track with positive or negative lifetime sign, leading to mostly symmetric IP distributions (as seen in Figure 6.7). A data augmentation procedure called *flipping* can be applied whereby the sign of track IPs (and that of secondary vertices) is multiplied by -1, without affecting the overall light-jets IP distributions [120]. The tagger evaluated on flipped inputs, the *flipped* tagger, has approximately equal performance for light-jets as the nominal tagger. However, for b -jets and c -jets

with real large IP tracks, the flipping will lead to large changes in their asymmetric IP distribution, with significantly fewer large IP tracks, causing the flipped tagger to be inefficient for identifying heavy flavor jets. Since RNNIP was a new low-level tagger, we wanted to ensure its design allowed a good flipped tagger definition. Compared to [93], the ordering of the tracks was from $|s_{d0}|$ to s_{d0} to preserve the lifetime sign information and enhance the flipping performance. Then the flipped RNNIP (RNNIPFlip) multiplies the IP signs of the s_{d0} and s_{z0} inputs by -1. Figure 6.33 shows the RNNIP discriminants of b -jets, c -jets and light-jets with nominal and flipped inputs. The dashed vertical lines represent the discriminant requirement for 85%, 77%, 70% and 60% b -jet efficiencies. The desired properties are found: the flipped distribution for light-jets is nearly unchanged, while there is a significant decrease in flipped b -jets and c -jets at high discriminant values. Using these distributions, the efficiencies of the different jet flavors as a function of the discriminants are examined in Figure 6.34. There is a large reduction in the efficiency for selecting b -jets and c -jets for a fixed light-jet rejection as desired.

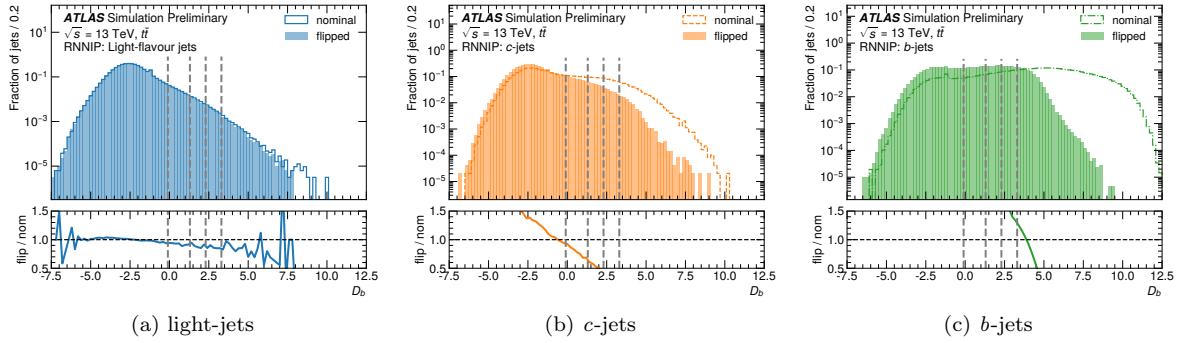


Figure 6.33: D_b discriminant distributions for the nominal and flipped taggers. The vertical dashed lines correspond to the discriminant requirements for 85%, 77%, 70%, and 60% inclusive b -jet efficiencies, corresponding to the efficiency benchmarks used at the analysis level. [5]

Given these promising results for the light-jet calibration, a corresponding DL1Flip tagger was defined using this RNNIP flip definition and the DL1r light-jet calibration was performed. In the Z +jets sample, the leading p_T jet in the event is the one considered for the b -tagging calibration, and a 2d fit with the pseudo-continuous b -tagging bins and the SV1 mass, where the SV1 mass helps constrain the flavor fractions. The light-jet distributions for the nominal and flipped DL1r tagger is shown in Figure 6.35(a) [4]. Again the nominal and flipped taggers have similar distributions, and the residual differences are included as a systematic in the calibration. The result of the calibration as the 70% WP is shown in Figure 6.35(b), and the SF is within 20% of the nominal value.

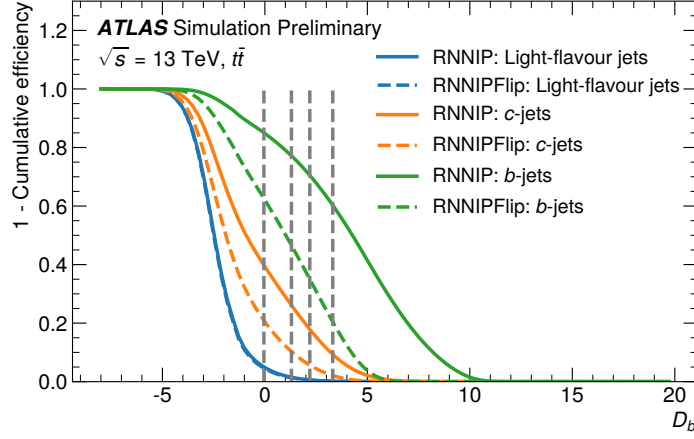


Figure 6.34: 1 - Cumulative efficiency as a function of b -tagging discriminant for RNNIP (a) and DIPS (b). In both cases, the performance remains nearly unchanged for light-flavour jets when comparing nominal and flipped taggers, while the b -jet and c -jet efficiencies drop. [5]

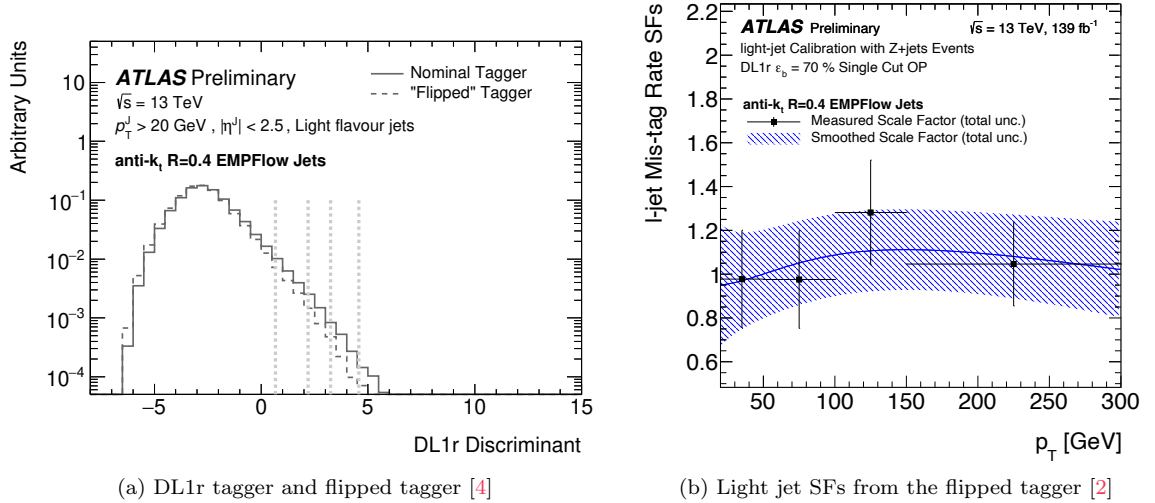


Figure 6.35: Performance of the Z +jets light jet calibration with the DL1rFlip tagger.

7

b-tagger R&D: DIPS

She raised one leg and gave me all her weight as I dipped her. She either trusted me or wanted to fall.

– John Green, *Paper Towns*

In this chapter, the Deep Impact Parameter Sets (DIPS) b -tagging algorithm is presented. It is based on the Deep Sets architecture [121], which is an alternative to the RNN b -tagging approach described in Chapter 6.3.2, but instead treats the tracks in the jet as an unordered, variable-sized *set* rather than as a *sequence*. This avoids the need to specify a track ordering, which for b -tagging was empirically driven rather than dictated by the problem. DIPS is demonstrated to be as performant as RNNIP but faster to train, which reduces turn-around time for optimization. As such, track selection criteria and extra track feature optimization studies are presented. Finally, diagnostic studies from the machine learning literature are included to demonstrate the characteristics from the b -quark fragmentation and hadronization process that the network has gleaned.

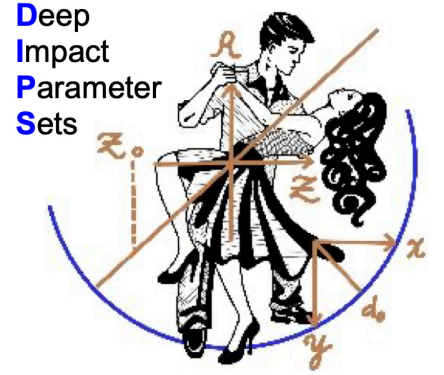


Figure 7.1: Logo for the DIPS algorithm. The couple is featured doing a dance move called a “dip” and the tan axes overlay the impact parameter definitions. Logo artist, Brendan Hartman.

7.1 DIPS Algorithm

The Energy / Particle Flow Network is first application of Deep Sets architecture to particle physics [122], and the motivation for this study. Let X_i be the vector representing the inputs associated with the i^{th} track in the jet. The Deep Set applies a neural network (NN) Φ to each track, sums over the tracks, and then applies additional processing on the summed representation with a feed-forward NN F , as described in Eq. 7.1,

$$\mathcal{O}(\{X_1, \dots, X_n\}) = F\left(\sum_{i=1}^n \Phi(X_i)\right), \quad (7.1)$$

where $\mathcal{O}(\{X_1, \dots, X_n\})$ represents the b -, c -, and light-jet class probabilities derived from the inputs for the n tracks in the jet. The architecture bifurcates the problem into operations over inputs and operations over sets, where the track-network Φ extracts the relevant track features, and the jet-network F accounts for the correlations between the tracks. The permutation invariance of the set is encoded with the permutation invariant sum operation, although other permutation invariant operations such as the max or average could be used as well.

DIPS offers the same advantages as RNNIP but encodes permutation invariance between the tracks in the jet, giving a more natural representation of the data and allowing the algorithm to be trained more efficiently with fewer parameters and less data [123]. In addition, Deep Sets offers a major additional advantage over RNNs in that the operation of processing the tracks in the jet with the Φ network can be easily parallelized. This allows training and evaluation to make significantly more efficient use of the GPUs compared to the non-parallelizable iterative processing of the RNN.

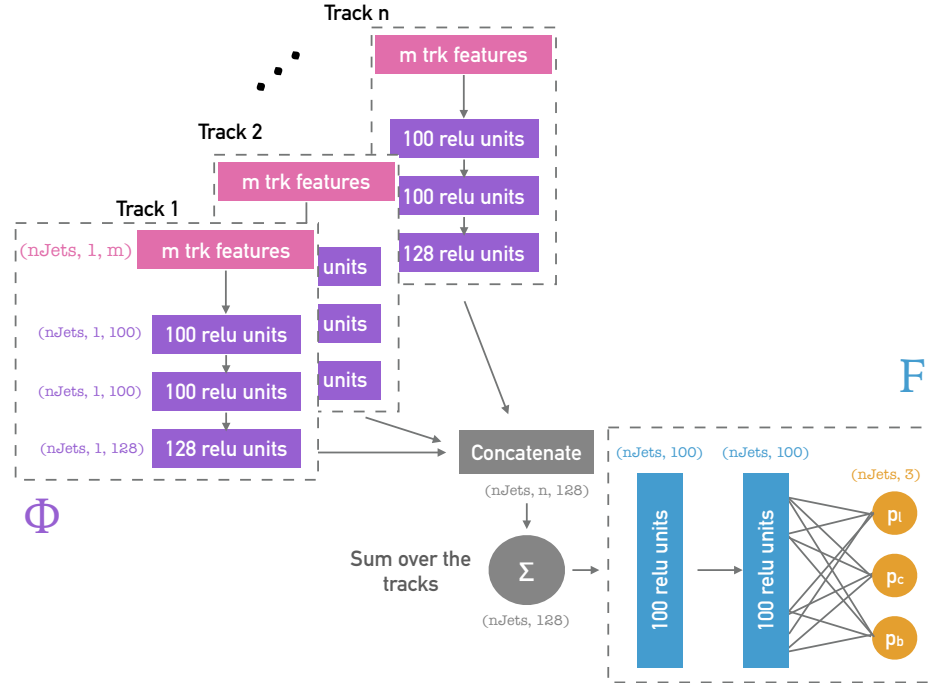


Figure 7.2: Architecture for the DIPS algorithm. The number of hidden units in the different neural network layers correspond to the final optimized architecture.

A simplified scheme of the DIPS architecture is shown in Figure 7.2, which is based on the architecture in [122]. A grid search over the hyperparameters including the number of layers in the Φ

and F networks, the number of nodes in the Φ and F networks, and the dimension of the track latent space revealed similar performance for several different choices of these hyperparameters. A batch normalization layer is included before the non-linearities, which is a regularization technique that normalizes the activations to have zero mean and unit variance [124]. This encourages the hidden representation to stay close to the turn-on of the non-linearities and learn more efficiently. We found batch norm was helpful for the DIPS b -tagging performance while dropout was not.

In this chapter, algorithms are just trained on $t\bar{t}$ simulated events (described in Section 6.1). For the baseline DIPS and RNNIP comparisons, the same input features (specified in Table 6.3) are used for both algorithms, and the same “nominal” track selection ($|d_0| < 1$ mm, $|z_0 \sin \theta| < 1.5$ mm, and $p_T > 1$ GeV) is applied. The tracks are ordered by decreasing s_{d0} , and the first 15 tracks are kept for training and evaluation. Figure 7.3 justifies that keeping a maximum of 15 tracks is a conservative choice for as this is on the plateau for the average track multiplicity in the jet.

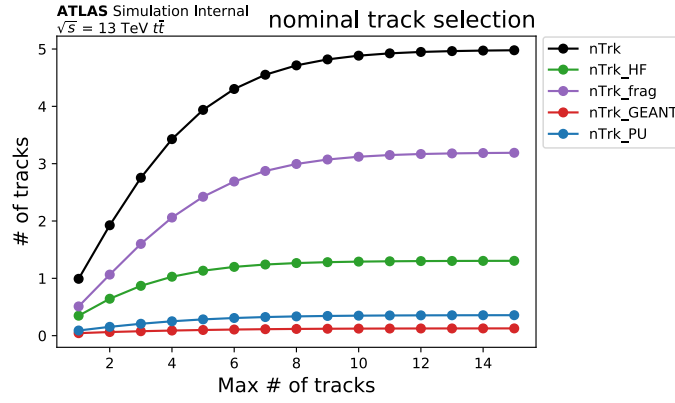


Figure 7.3: The track multiplicity (black) in the jet given the cut on the maximum number of tracks cut. b , c , and light jets are inclusively considered, with the nominal track selection. The track multiplicity by the track origin is also shown for heavy flavor tracks (green), fragmentation tracks (purple), tracks from material interactions (red) and tracks from other vertices (blue).

The RNNIP and DIPS trainings are performed with 3 million jets, with 20% of these jets held out as a validation set to determine when to stop the training. The training is terminated when the validation loss has not improved in the last ten epochs, and the model with the best validation loss is selected. Both the RNNIP and DIPS architectures are implemented in Keras [125] and trained with the TensorFlow backend [126]. Algorithms were trained with the Adam optimizer [112] with a learning rate of 10^{-3} and a batch size of 256. The performance metrics shown in the following sections use a statistically independent dataset of 3 million jets.

7.2 Baseline performance

Each model is trained five times and for a given b -jet efficiency, the mean of the rejections is used as the nominal value and the standard deviation of the rejections gives the width of the curve. This ensemble of trainings assesses the predictive uncertainty of machine learning-based algorithms [127].

The ROC curves for b -jet efficiency versus light-flavor jet rejection and for b -jet efficiency versus c -jet rejection of the DIPS and RNNIP algorithms are shown in Figure 7.4. The DIPS algorithm provides up to a 15% additional light-flavor jet rejection and a 5% additional c -jet rejection at a given b -jet efficiency over the RNNIP algorithm.

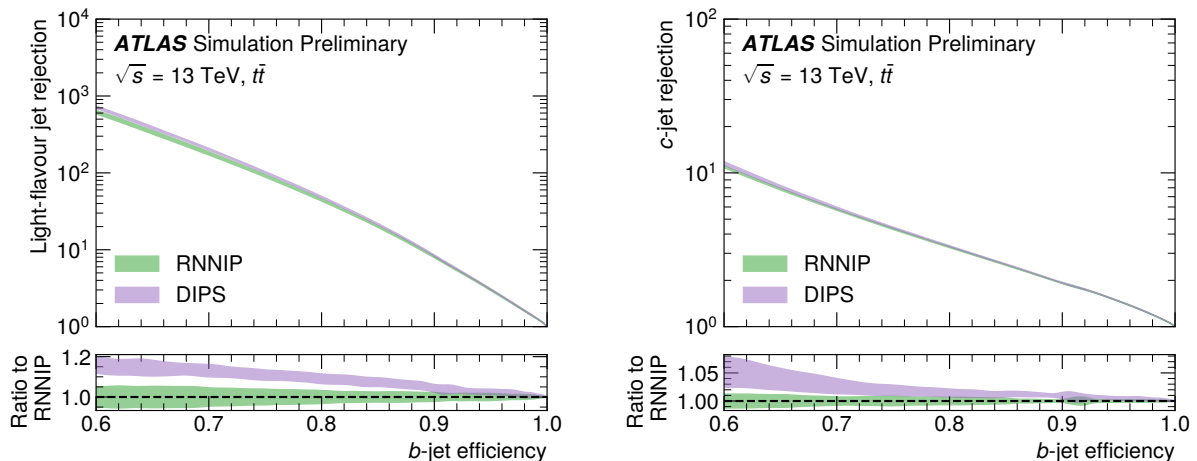


Figure 7.4: Light-flavour jet rejection as a function of b -jet efficiency (left) and c -jet rejection as a function b -jet efficiency (right) of the RNNIP (green) and DIPS (purple) algorithms. The central curves and error bands show the mean and standard deviation, respectively, of the rejection at each b -jet efficiency for 5 trainings. The ratios are computed with respect to the RNNIP ROC curve. [5]

This similar performance comes with a significant decrease in training and evaluation time. The training time limits the ability to perform optimization tests that compare model variants, while the evaluation time impacts the ATLAS reconstruction time for deploying the model at scale, and the ability to use such algorithms in low-latency environments such as the trigger. The DIPS and RNNIP models with similar numbers of parameters are compared in terms of their speed of training and evaluation in Table 7.1. Training comparisons are done on an NVIDIA 2080 Ti GPU, while evaluation comparisons are performed on an NVIDIA Titan X GPU. Five versions of each model are trained and evaluated, and the mean and standard deviation of the training and evaluation time is reported. There is a significant speed-up of more than a factor of 2 for the DIPS algorithm compared to RNNIP. For the evaluation times, DIPS is nearly a factor of 4 faster than RNNIP.

Model	Parameters	Time / epoch [s]	GPU eval time [s]	CPU eval time [s]
RNNIP	47k	241 ± 14	170 ± 2	685 ± 84
DIPS	49k	78 ± 4	46 ± 2	206 ± 98

Table 7.1: Timing metrics for trainings on an Nvidia 2080 Ti GPU and evaluations on an NVIDIA Titan X GPU. The nominal value denotes the mean of five independent trainings, while the error bar is the standard deviation. These metrics are for the inference and evaluation are for the full test datasets of 3 million jets each [5].

7.2.1 Interpretability

To explore what DIPS is learning in the correlations between features that aids the classification performance, the average saliency map for b -jets with 8 associated tracks and failing a threshold corresponding to 77% b -tagging efficiency is shown in Figure 7.5. The saliency map is computed as

$$\frac{\partial D_b}{\partial x_{ik}} = \frac{1}{N} \sum_{j=1}^N \frac{\partial D_b^{(j)}}{\partial x_{ik}^{(j)}}, \quad (7.2)$$

and is the gradient of the discriminant value $D_b^{(j)}$ with respect to each track feature input $x_{ik}^{(j)}$, averaged over jets (j) in a sample of N jets [128]. The feature inputs normalized to zero mean and unit variance in the training use the same preprocessing in this evaluation. The saliency map is a first-order view of the discriminant’s response to the inputs. Figure 7.5 thus shows how this sample of b -jets that failed tagging could be modified to make them more b -jet-like. There are reasonably strong positive gradients for the significances (s_{d_0} and s_{z_0}) extending up to 5 tracks, which is the average number of charged particle tracks in a b -hadron decay. Beyond 5 tracks, the gradients for all the features are either nearly zero or negative, indicating that either these tracks provide no further information, or that tracks with large IPs are more indicative of background. In addition, DIPS is highly sensitive to the $\log p_T^{frac}$ and $\log \Delta R$ of the leading s_{d_0} track, which is consistent with the harder fragmentation of b -quarks compared to light and charm jets. Interestingly, this strong correlation with $\log p_T^{frac}$ and $\log \Delta R$ for the highest s_{d_0} track also indicates that simply enlarging the IPs of a track in a jet would not directly lead to a jet passing a tagging threshold, as the track must also be consistent with the kinematic expectations from b -jet fragmentation. The gradients for the shared and split hits of the high s_{d_0} tracks are strongly negative since tracks formed from random combinations of hits are more likely in highly dense environments. Also, the correlation with the overall number of hits in the innermost pixel layers, IBL and PIX1, is positive but small. Such features are of high importance to the estimate of the IP and IP resolution. However such information is also encapsulated in the IP significance features which are strongly correlated with the discriminant. We suspect the IBL and PIX1 gradients are relatively small due to the discriminator heavily relying on the IP significance for the first-order estimate of the track’s quality.

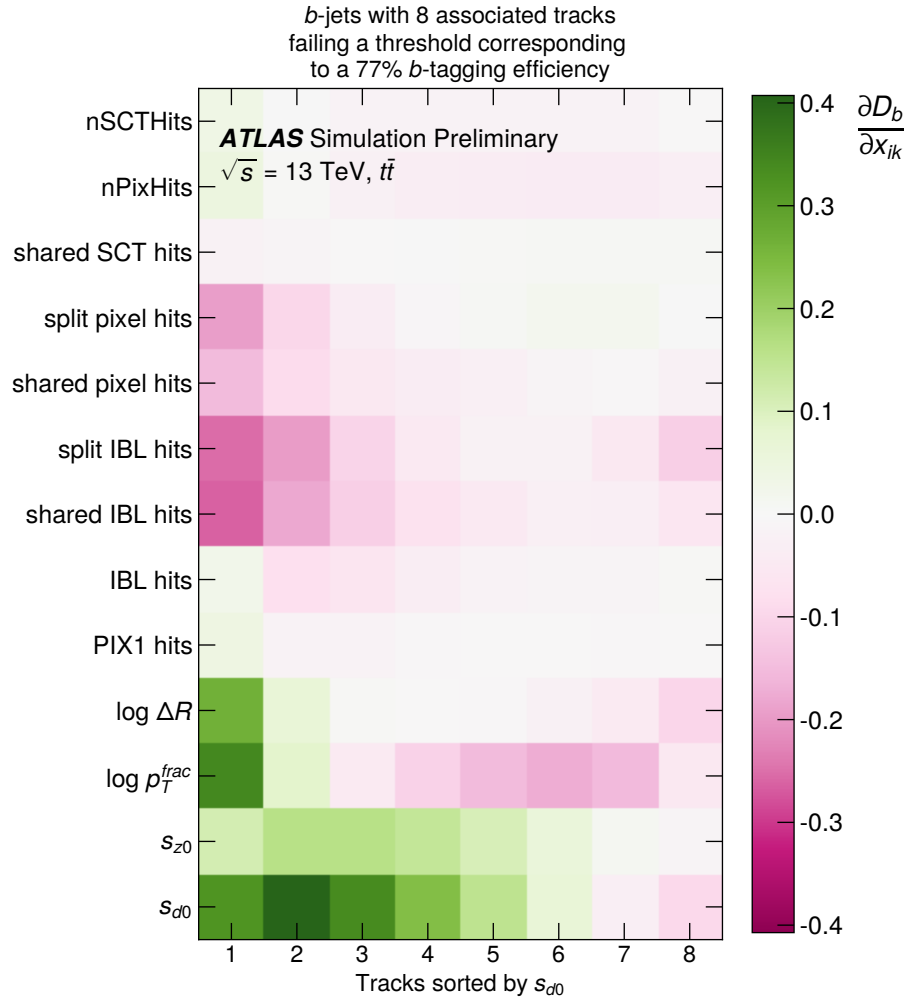


Figure 7.5: Saliency map for *b*-jets with 8 tracks. The track features are shown on the *y*-axis, the tracks (ordered by s_{d0}) are listed on the *x*-axis. The colors in each pixel represent the gradient defined in Equation 7.2 [5].

7.2.2 Calibratability

Chapter 6.5 emphasized the importance of designing new taggers with a valid “flipped definition” for the light jet calibration. We designed a “DIPSFlip” tagger similar to what was done for RNNIPFlip by modifying the inputs to multiply the IP significances by -1. The same metrics are shown as were considered before for RNNIPFlip. Figure 7.6 shows the discriminants for DIPS and DIPSFlip for the light, *c*, and *b*-jet distributions. These distributions are again largely the same for light jets, while the performance for *c*, and *b*-jets are reduced (the distribution is shifted to the lower D_b values), as desired for the flipped tagger definition. This is further exemplified in the integral of these

distributions in Figure 7.7. DIPSFlip has similar performance for light jets, but lower efficiencies for c and b -jets. Since the DIPS tagger has a permutation invariance over the tracks, while RNNIP uses the lifetime signed s_{d0} for the track ordering, DIPSFlip is not quite as performant as RNNIPFlip: the heavy flavor jet efficiency is slightly larger for DIPSFlip compared to RNNIPFlip (Figure 6.34). However, we still believe that the DIPS performance metrics are still promising for calibrating the high-level tagger with the DIPS with the same “flipped tagger” methodology.

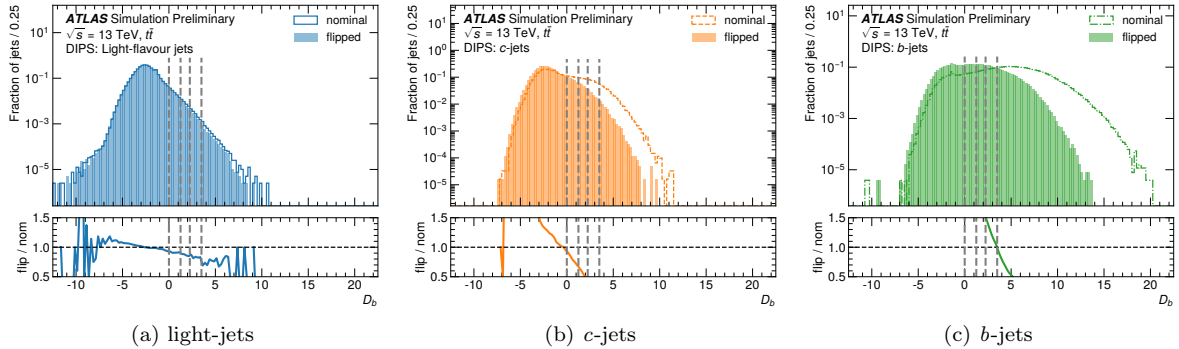


Figure 7.6: D_b discriminant distributions for the nominal and flipped DIPS taggers. The vertical dashed lines correspond to the discriminant requirements for 85%, 77%, 70%, and 60% inclusive b -jet efficiencies, corresponding to the efficiency benchmarks used at analysis level. [5]

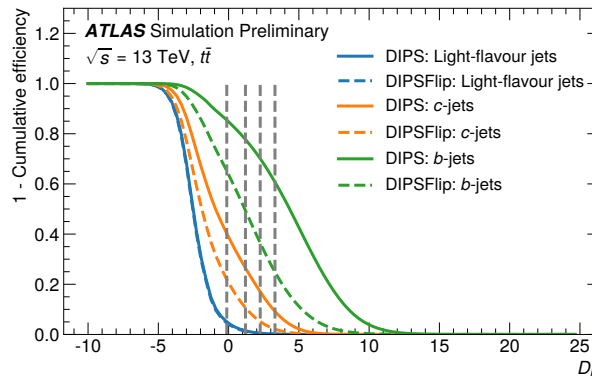


Figure 7.7: 1 - Cumulative efficiency as a function of b -tagging discriminant for DIPS. The performance remains nearly unchanged for light-flavor jets when comparing nominal and flipped taggers, while the b -jet and c -jet efficiencies drop. [5]

7.3 Track Selection Optimization

A major benefit of the reduced training time for DIPS is that it facilitates optimization studies which require retraining the algorithm with many configurations. Two optimizations are presented here: (1) loosening the input track selection, and (2) providing additional features per track.

The DIPS implementation described so far uses the same *nominal* track selection ($|d_0| < 1$ mm, $|z_0 \sin \theta| < 1.5$ mm, $p_T > 1$ GeV) as the IP3D and RNNIP algorithms. This high purity selection keeps the number of mismeasured and pile-up tracks low, as the IP3D algorithm can be sensitive to such tracks, and Naïve Bayes models perform worse as the number of inputs increases [129].

However, the nominal selection removes some of the key tracks from heavy flavor decays vital for classification. With the greater expressivity of the DIPS NN compared to the IP3D model, DIPS can learn which tracks are useful for tagging to be less sensitive to poor-quality tracks. A *loose* selection is examined, defined as $|d_0| < 3.5$ mm, $|z_0 \sin \theta| < 5$ mm, $p_T > 0.5$ GeV. DIPS with the *loose* selection examines up to the 25 highest s_{d_0} tracks, and Figure 7.8 justifies that keeping a max of 25 tracks is a conservative choice as this is well on the plateau for the average number of tracks in the jet.

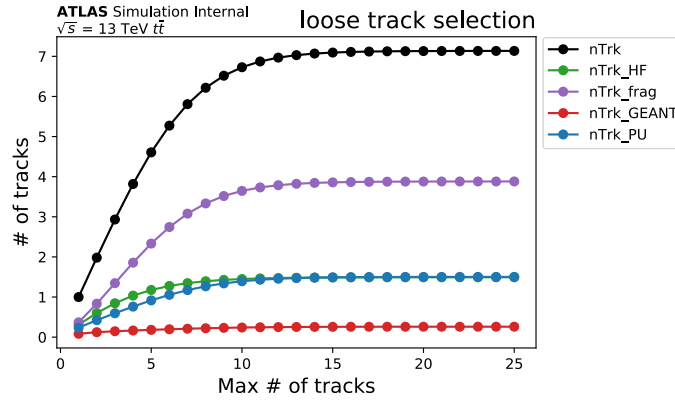


Figure 7.8: The track multiplicity (black) in the jet given the cut on the maximum number of tracks cut. b , c , and light jets are inclusively considered, with the nominal track selection. The track multiplicity by the track origin is also shown for heavy flavor tracks (green), fragmentation tracks (purple), tracks from material interactions (red) and tracks from other vertices (blue).

The average number of tracks per jet in $t\bar{t}$ is shown in Table 7.2 for the *nominal* and *loose* selections, and is shown separately for each jet flavor. The total number of tracks (n_{trk}), the number of tracks from heavy flavor decays (n_{trk}^{HF}), the number of tracks from hadronization, excluding those from heavy flavor decays (n_{trk}^{hadr}), and the number of tracks from mismeasurement, material interactions, and pile-up (n_{trk}^{other}), are compared. The highlighted block emphasizes that the *loose* selection increases the average number of heavy flavor tracks per b -jet by $\approx 15\%$ over the *nominal* selection. There is also a 6% increase in the number of heavy flavor tracks in c -jets. The *loose*

selection did increase the number of fragmentation and other tracks per jet. However, the higher heavy flavor track efficiency was more impactful. As can be seen in the ROC curves in Figure 7.9, DIPS with the *loose* selection (shown in cyan) outperforms the nominal DIPS (shown in purple) by up to $\approx 40\%$ for light-jet and charm jet rejection.

Jet Flavour	Track selection	n_{trk}	n_{trk}^{HF}	n_{trk}^{hadr}	n_{trk}^{other}
<i>b</i> -jets	<i>nominal</i>	5.9 ± 2.7	3.4 ± 1.8	2.0 ± 1.9	0.4 ± 0.8
	<i>loose</i>	8.1 ± 3.2	3.9 ± 1.8	2.5 ± 2.1	1.7 ± 1.7
<i>c</i> -jets	<i>nominal</i>	5.1 ± 2.5	1.7 ± 1.0	2.9 ± 2.2	0.4 ± 0.8
	<i>loose</i>	7.1 ± 3.1	1.8 ± 1.0	3.6 ± 2.4	1.7 ± 1.7
Light-flavour jets	<i>nominal</i>	4.6 ± 2.6	-	4.1 ± 2.5	0.5 ± 0.9
	<i>loose</i>	6.8 ± 3.3	-	5.0 ± 2.7	1.8 ± 2.0

Table 7.2: The average per jet total number of tracks (n_{trk}), the number of tracks from heavy flavour decays (n_{trk}^{HF}), the number of tracks from hadronisation, excluding those from heavy flavour decays (n_{trk}^{hadr}), and the number of tracks from mismeasurement, material interactions, and pile-up (n_{trk}^{other}), are shown for the *nominal* and *loose* selections for each jet flavor [5].

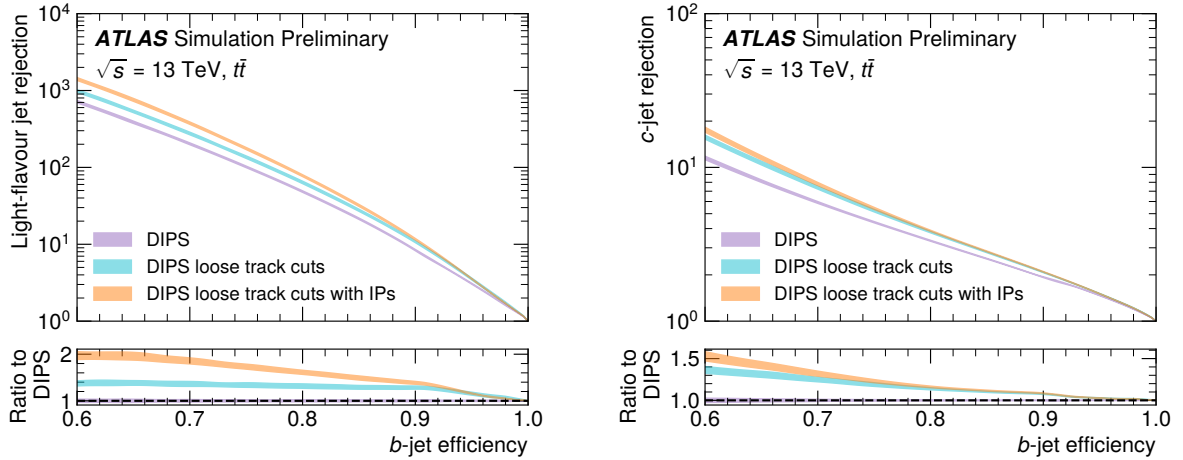


Figure 7.9: Light-flavour jet rejection as a function of *b*-jet efficiency (left) and *c*-jet rejection as a function of *b*-jet efficiency (right) of the nominal DIPS setup, DIPS with *loose* track selection, and Optimized DIPS with the *loose* track selection and additional IP inputs. The central curves and error bands show the mean and standard deviation, respectively, of the rejection at each *b*-jet efficiency for 5 trainings. The ratios are computed with respect to the DIPS ROC curve. [5]

Beyond the *loose* selection, adding more per-track features is also examined, namely the impact parameters d_0 and $z_0 \sin \theta$. DIPS with the additional features and the *loose* track selection is called “Optimized DIPS”, and is shown in orange in the ROC curves in Figure 7.9. It is compared to the

RNNIP and nominal DIPS references. The Optimized DIPS outperforms the nominal DIPS by up to a factor of 2 in light-jet rejection and a factor of 1.5 in the c -jet rejection.

The b -jet efficiency as well as the c -jet and light-flavour jet rejections versus jet p_T and η are shown in Figure 7.10, for the RNNIP, DIPS, and Optimized DIPS algorithms. The behavior of DIPS and RNNIP are nearly the same across the p_T and η range, with DIPS providing a slightly higher light-jet rejection. The Optimized DIPS delivers a factor of 1.5 to 2.5 additional light-jet rejection and up to $\approx 33\%$ additional c -jet rejection. Loosening the track requirements for Optimized DIPS could potentially have the drawback of increasing the performance dependency on pile-up. We therefore check the b -jet efficiency, c -jet and light-jet rejection as a function of the average number of proton-proton collisions per bunch crossing $\langle\mu\rangle$, also shown in Figure 7.10. The Optimized DIPS performance dependency on $\langle\mu\rangle$ is not found to be significantly stronger than the baseline DIPS or RNNIP.

A comparison of the c -jet and light-jet rejections as a function of p_T and η for the DIPS, RNNIP, and Optimized DIPS algorithms with a flat 77% b -tagging efficiency can be seen in Figure 7.11. Optimized DIPS shows more than a factor of 2 increase in light-jet rejection and up to $\approx 30\%$ additional charm jet rejection compared to RNNIP, for jets with p_T between 50 and 300 GeV.

7.4 A look to the future

Since the studies presented in [5], the DIPS has been retrained with a seven-fold increase in the training set size. With this more sophisticated training regime, the improvements compared to RNNIP became even more dramatic, giving more than a factor of three improvement for the light-jet rejection and up to two times better for charm jet rejection in $t\bar{t}$ evaluations. When this improved DIPS training was propagated to a high-level DL1d tagger¹, these DIPS improvements led to an **up to 50% improvement** in DL1d relative to the DL1r presented in Chapter 6.4.1. The b -tagging performance has not saturated yet, as a Graph Neural Network approach, called GN1, is showing very promising results [130]. This is similar to the DIPS algorithm as the tracks in the jet are modeled as a set, but the architecture additionally allows the flexibility to have “edge features” for the pair-wise connections between all of the tracks, conceptually similar to how the vertexing algorithms always start with all pairs of two track vertices. GN1 adds auxiliary terms to the loss to (1) determine if tracks belong in the same vertex, and (2) classify the origin of the tracks. GN1 uses all of the input features that we found useful for the Optimized DIPS, but add additional inputs, such as the rest of the perigee inputs. The result is a 2x improvement in both light and charm rejection. GN1 will be the new b -tagger for Run 3, but it’s exciting for DIPS to have been an optimization along the way for such strong b -tagging results.

¹The “d” in DL1d refers to the DIPS input for the high-level tagger

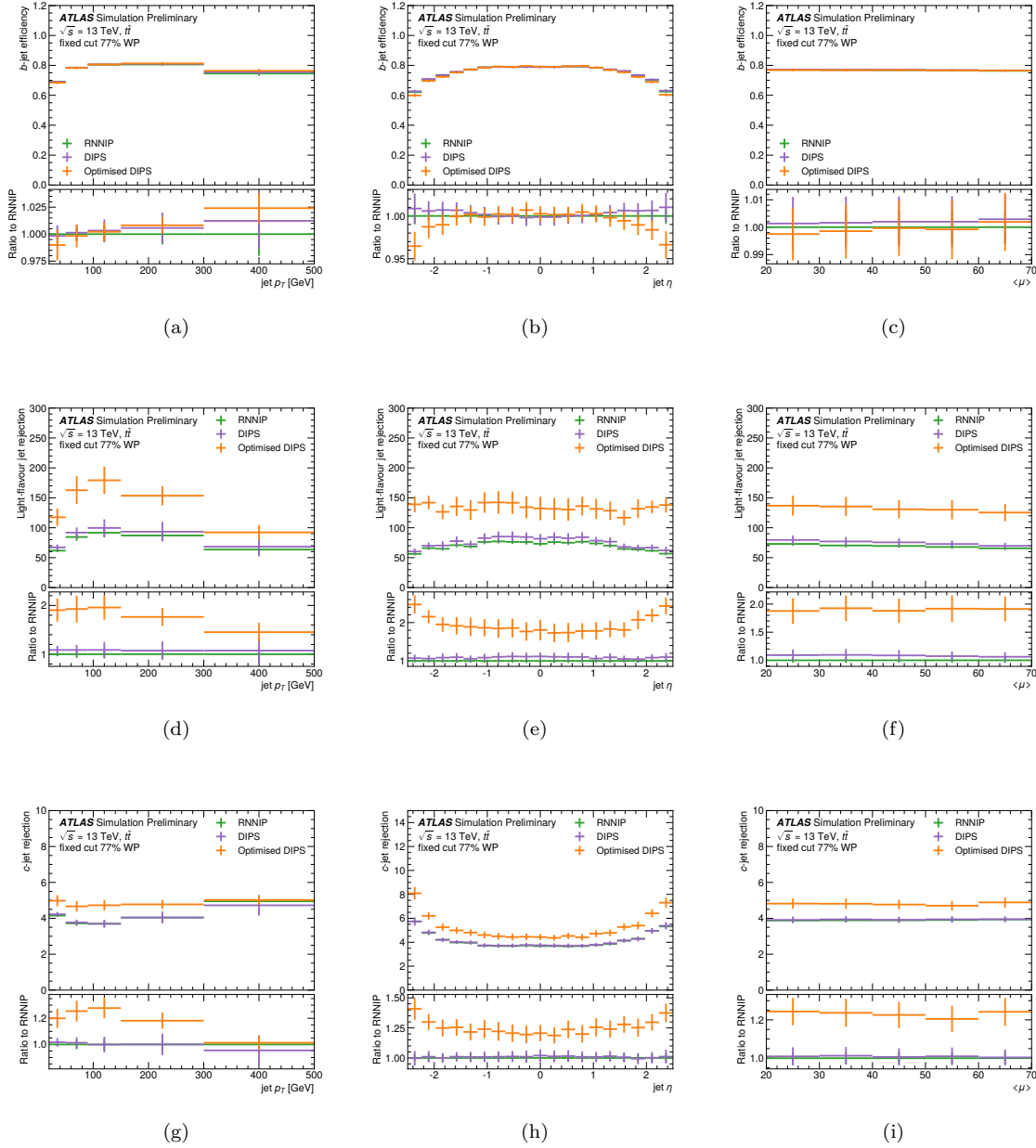


Figure 7.10: Performance plots using a fixed cut with 77% *b*-jet efficiency. Plots (a), (b) and (c) show the *b*-jet efficiency as a function of jet p_T , η and average number of proton-proton collisions per bunch crossing $\langle \mu \rangle$. Plots (d), (e) and (f) show the light-jet rejection as a function of the same quantities, while plots (g), (h) and (i) show the *c*-jet rejection. [5]

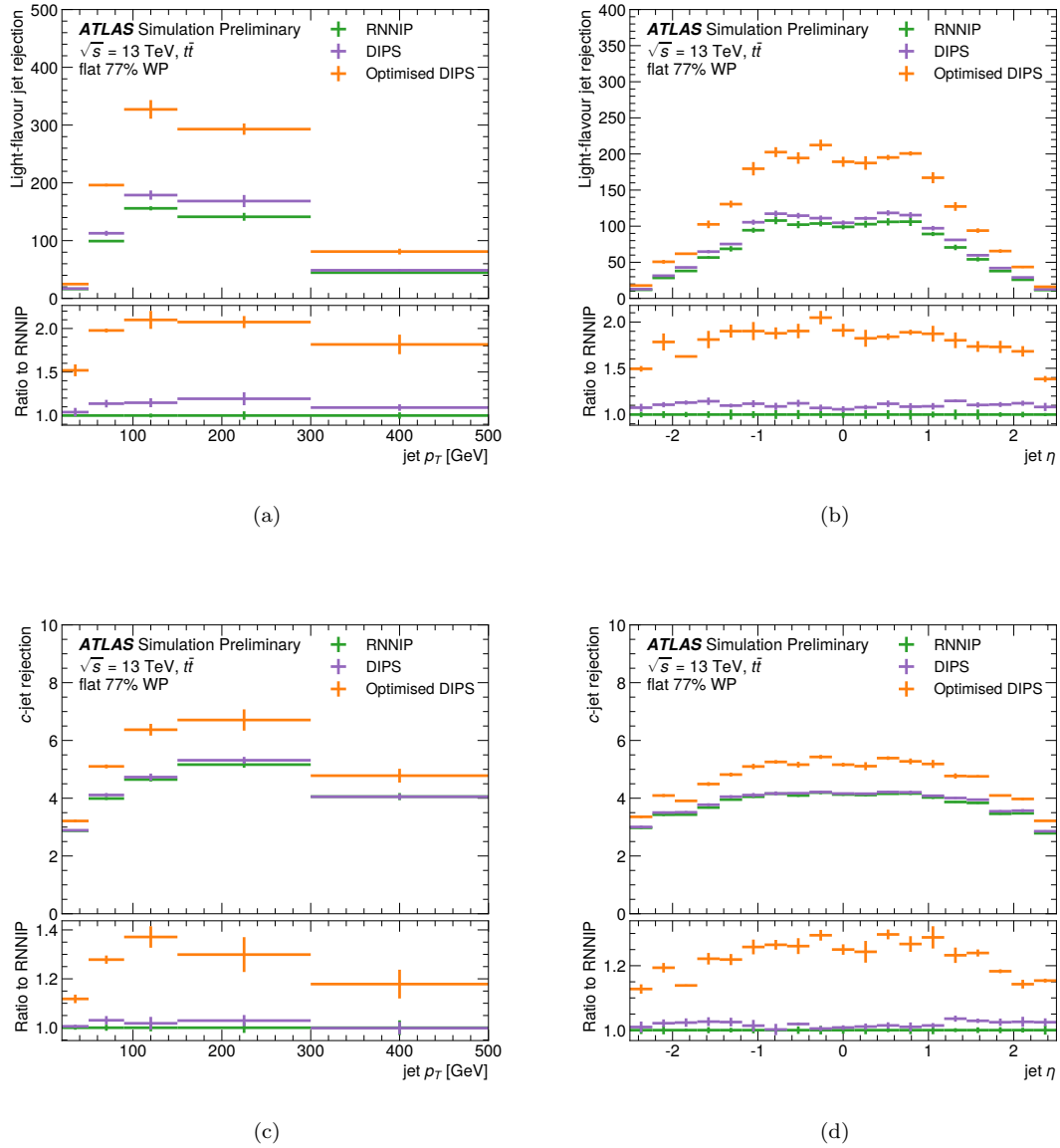


Figure 7.11: Performance plots using a requirement where the b -jet efficiency is 77% in each bin. Plots (a) and (b) show the light-jet rejection as a function of jet p_T and η , while plots (c) and (d) show the c -jet rejection as a function of the same quantities. [5]

Part III

$HH \rightarrow 4b$ analysis

This part reviews the non-resonant $HH \rightarrow 4b$ analysis – the key physics contribution of this thesis. In Chapter 8 we begin by describing the statistical analysis framework that will be used to interpret the results. Chapter 9 then motivates HH physics as a way to probe the form of the Higgs potential, and describes the signal parametrization with κ framework and other effective field theory frameworks. Chapter 10 describes the selections of the ATLAS $HH \rightarrow 4b$ analysis, and the final discriminant used for hypothesis testing. Chapter 11 elaborates on the fully data-driven background estimate, the systematics assessed for the method, and the suite of validation studies performed to verify the procedure. Chapter 12 concludes with the $HH \rightarrow 4b$ results, also including the subsequent impact in the combination with the other main channels.

Statistical techniques

The theory of probabilities is at bottom nothing but common sense reduced to calculus; it enables us to appreciate with exactness that which accurate minds feel with a sort of instinct for which oftentimes they are unable to account.

– Pierre Simon Laplace

In a physics analysis, we search for a signal amidst background processes. We formalize this search through *hypothesis testing* by defining both null (H_0) and alternative (H_1) hypotheses. In a hypothesis test, all we can do is reject the null hypothesis in favor of the alternative [131]. The *p-value* is the probability of the observed data if the null hypothesis were true, and we reject the null if this *p-value* is less than a pre-defined threshold, α . Figure 8.1 gives a simple signal and background model, and shows the two different hypothesis testing scenarios commonly considered. For a discovery (i.e, the Higgs discovery), the null hypothesis is that there is no signal, and the

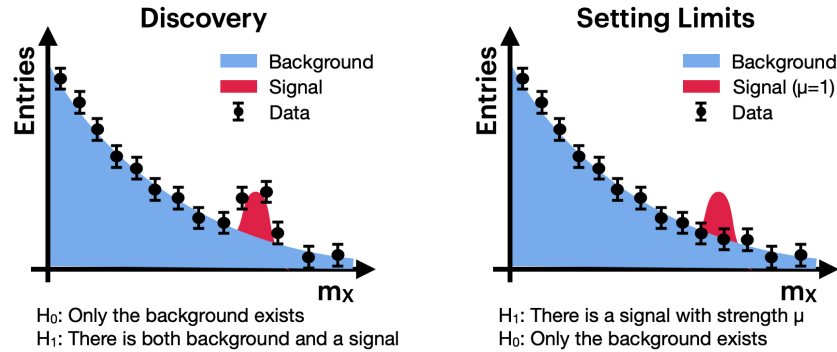


Figure 8.1: Hypothesis testing cases used in particle physics. Shown are histograms for a background and signal models for an invariant mass m_X . The left image shows a data distribution for observing the signal processes, while the right shows a data distribution for setting limits on the signal strength.

alternative is that the signal exists. The standard in particle physics to claim a disclaim a discovery

is a “5 σ ” significance, meaning that the probability under the Gaussian more than 5σ away from the mean, which corresponds to $\alpha = 2.7 \cdot 10^{-7}$. When the signal is too small to see, the null hypothesis is that the signal *exists* with a signal strength (or normalization factor) μ , and we test how large μ can be before the probability of the data is less than $\alpha = 0.05$, which is a less strict threshold for limit setting than discovery. This chapter outlines the statistical methodology used for setting the limits for the physics results presented in Chapter 12.

8.1 The likelihood

The central quantity for these statistical tests is the *likelihood* which describes the probability of the data under the null hypothesis. For a counting experiment, the probability of having a n events if the expectation was λ is given by the Poisson distribution:

$$\text{Poisson}(n|\lambda) = \frac{\lambda^n}{n!} e^{-\lambda}.$$

The expected number of events, λ , is the sum of the signal (s) and the background (b) yields: $\lambda = \mathbb{E}[n] = \mu \cdot s(\theta) + b(\theta)$. where μ is the signal strength (or *parameter of interest*, PoI) and θ are *nuisance parameters* (NPs), a part of the model we must estimate but is not of primary interest. For example, SFs that correct the b -tagging efficiency from simulation to data have an associated uncertainty that enters the likelihood by a NP that can affect the signal and background models.

The power of our analysis can be increased by binning (or categorizing) with a variable that offers additional signal-versus-background discrimination power so that $\mathbf{n}, \mathbf{s}, \mathbf{b} \in \mathbb{R}^N$ representing values in each of the N bins. Each histogram bin is an independent counting experiment with $\mathbb{E}[n_i] = \mu s_i(\theta) + b_i(\theta)$. The likelihood of the full histogram is thus the product of the Poisson probabilities in each bin. The NPs are often constrained by auxiliary measurements with $\mathbb{E}[m_i] = u_i(\theta)$ for $i = 1, \dots, M$, where $\mathbf{u}(\theta)$ is a model in a control region which also depends on the NPs, and \mathbf{m} are the observed data yields in the M entries of the control region histogram. These constraints are multiplied in as additional Poisson probabilities so that the full likelihood becomes:

$$\mathcal{L}(\mu, \theta) = \prod_{i=1}^N \frac{\mu s_i(\theta) + b_i(\theta)}{n_i!} e^{-(\mu s_i(\theta) + b_i(\theta))} \prod_{j=1}^M \frac{u_j(\theta)^{m_j}}{m_j!} e^{-u_j(\theta)} \quad (8.1)$$

8.2 Test statistic

To set up a hypothesis test, we use the **profile likelihood ratio**:

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\theta}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\theta})}, \quad (8.2)$$

where $\hat{\theta}(\mu)$ are the “profiled” nuisance parameters that maximize the likelihood for the signal strength μ , and the $\hat{\mu}$ and $\hat{\theta}$ are the unconstrained fitted values for the maximum likelihood estimates. This $\hat{\mu}$ is the best-fit signal strength. Note, $\lambda(\mu) \leq 1$, and small values for $\lambda(\mu)$ correspond to unlikely signal strengths. As multiplying small numbers can be numerically unstable – the log of the profile likelihood ratio is used to define a **test statistic**:

$$t_\mu = -2 \log \lambda(\mu), \quad (8.3)$$

where now larger values of t_μ are the less likely signal strengths. The distribution over the best-fit signal strength can be described by the probability distribution function (pdf) $f(t_\mu|\mu)$, and t_{obs} is the test-statistic for the observed data. Then we can calculate the p -value for this observation as:

$$p_\mu = \int_{t_{obs}}^{\infty} f(t_\mu|\mu) dt_\mu.$$

For setting upper limits, we use a one-sided test statistic q_μ :

$$q_\mu = \begin{cases} -2 \ln \frac{\mathcal{L}(\mu, \hat{\theta}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\theta})} & \hat{\mu} \leq \mu, \\ 0 & \hat{\mu} > \mu. \end{cases} \quad (8.4)$$

where the one-sided means only $\hat{\mu} < \mu$ are considered incompatible with the null hypothesis that a signal with strength μ exists. The p -value for this test statistic is defined similarly as:

$$p_\mu = \int_{q_{obs}}^{\infty} f(q_\mu|\mu) dq_\mu. \quad (8.5)$$

Additionally, we often want to consider the distribution of the test statistic q_μ under a different signal hypothesis μ' with the pdf $f(q_\mu|\mu')$. For example, the pdf $f(q_\mu|0)$ is used to find the upper limits for the signal strength under the hypothesis that no signal exists.

8.2.1 Asymptotic approximation

The p -value can not, in general, be calculated analytically. They can be estimated using Monte Carlo techniques. This involves (1) sampling the \vec{n} and \vec{m} from the Poisson probabilities defining the likelihood, (2) for each sample, calculating $\hat{\mu}$, $\hat{\theta}$, $\hat{\theta}(\mu)$, and then (3) using these profiled the parameters to construct the test statistic from Eq. B.1. The histogram of the q_μ samples approximates the pdf $f(q_\mu|\mu)$. This brute-force approach can be very time-consuming, especially as for hypothesis testing, we care about accurately modeling the tails of the distributions to calculate the probability under the null hypothesis (as given in Eq. 8.5).

To avoid this time-consuming procedure – [132] identified closed form solutions valid as the

number of events K is large enough. This comes from Wald's Theorem [133], which says that in the limit of large N , the profile likelihood ratio test statistic can be written as:

$$t_\mu = \frac{(\mu - \hat{\mu})^2}{\sigma^2} + \mathcal{O}\left(\frac{1}{\sqrt{K}}\right)$$

with $\hat{\mu}$ distributed normally with mean μ' and variance σ^2 . We use this asymptotics approximation in the $HH \rightarrow 4b$ analysis (presented in Chapter 10). The expression for $f(q_\mu|\mu')$ is given in Appendix B, along with a toy study from [132] comparing with the analytic pdf with the sampling approach.

8.3 Setting limits: CL_s

Hypothesis testing directly with the probability under the null (Eq. 8.5) can lead to pathological cases when we observe fewer events than expected. The CL_s modifies this p -value to avoid these pathologies [134]. The probability under the null, p_{s+b} is the same as before, but we additionally calculate p_b , the probability the data could be *less extreme* than the observed q_μ if there were no signal ($\mu = 0$).

Figure 8.2 visualizes these integrals defined below for $q_\mu = q_1$:

$$p_{s+b} = \int_{q_{obs}}^{\infty} f(q_\mu|\mu' = \mu) dq_\mu$$

$$p_b = \int_{-\infty}^{q_{obs}} f(q_\mu|\mu' = 0) dq_\mu.$$

CL_s normalizes p_{s+b} by p_b :

$$CL_s = \frac{p_{s+b}}{p_b}$$

We reject the null if $CL_s < \alpha$, and Appendix B.2 shows how this avoids the pathologies that can come when using p_{s+b} directly. Since $p_b < 1$, $CL_s > p_{s+b}$, CL_s sets more conservative upper limits than p_{s+b} – but is a convention used across LHC analyses.

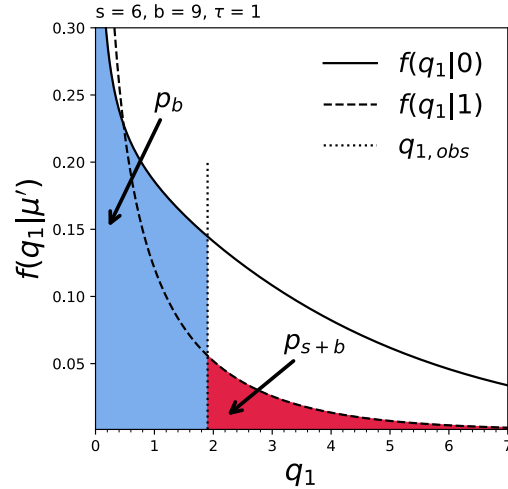


Figure 8.2: The solid and dashed lines show the pdfs for q_1 under the $\mu' = 0$ and $\mu' = 1$ hypotheses, while the dotted line shows an observed q_1 . p_{s+b} is the probability under $f(q_1|1)$ shaded to the right of the observed q_1 , while p_b is the probability under $f(q_1|0)$ to the left of the observed q_1 . Then $CL_s = p_{s+b}/p_b$.

Expected limits

Before unblinding the analysis and looking at the observed data, we calculate expected sensitivity based on the distribution over $\hat{\mu}$. Figure 8.3 illustrates this procedure which is summarized below.

1. The colored solid lines in Figure 8.3(a) indicate the $f(q_\mu|0)$ pdfs for $\mu = 0.7, 1$, and 1.5 . $q_{\mu A}$ is taken as the median of the $f(q_\mu|0)$ distribution, so $p_b = 0.5$ by definition. As μ increases, large q_μ are less compatible with the $\mu' = 0$ hypothesis, and $q_{\mu A}$ shifts to the right as μ increases.
2. Then $p_{s+b} = \int_{q_{\mu A}}^{\infty} f(q_\mu|\mu') = \mu dq_\mu$. $f(q_\mu|\mu)$ is shown in the black dashed line, and does not depend on μ , but the μ dependence p_{s+b} enters in the lower limit on this integral.
3. Then we can calculate the $CL_s = \frac{p_{s+b}}{p_b}$ for each μ , shown by the colored dots in Figure 8.3(b).
4. Steps (1) – (3) are repeated for a range of μ , and the black dashed line of Figure 8.3(b) shows the result of this μ scan. The expected 95% Confidence Level upper limit (95% CL limit) on μ is when $CL_s = 0.05$, and indicated by the “X”.
5. The distribution over $\hat{\mu}$ gives a variation over the μ s which we use to cite an uncertainty on the 95% CL μ_{exp} upper limit

The prescription for the observed upper limits on μ is the same, except the $q_{\mu,obs}$ test statistic is used for the CL_s calculation.

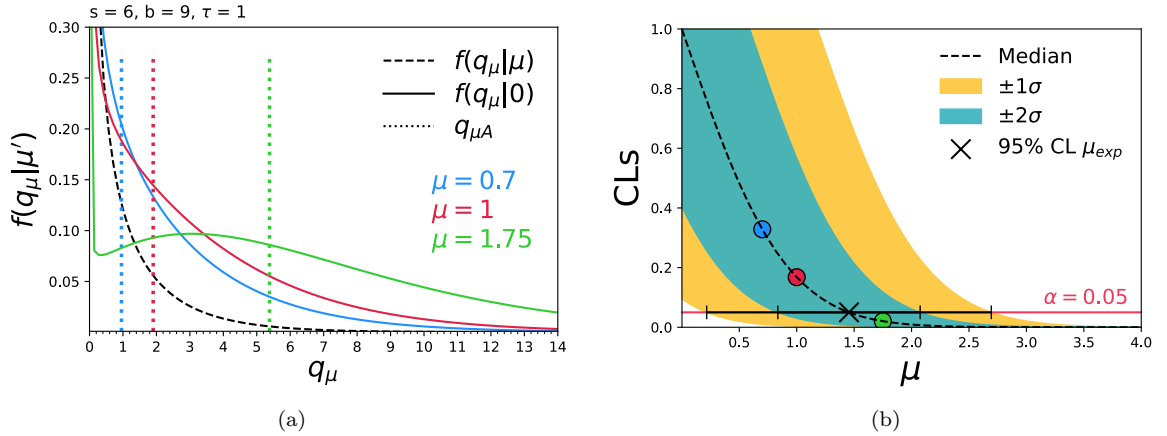


Figure 8.3: On the left are shown the the $f(q_\mu|\mu)$, $f(q_\mu|0)$ and $q_{\mu A}$ for $\mu = 0.7$ (blue), $\mu = 1$ (red), and $\mu = 1.5$ (green). The colored dots in on the right show the CL_s for each of these μ s. The expected 95% CL upper limit on μ is indicated by the “X”. The turquoise (yellow) bands show the $\pm 1\sigma$ (2σ) errors, and the the black error bars indicate the 95% CL μ_{exp} limit uncertainty.

8.4 Model implementation for analyses

We consider a few different types of NPs which we call the analysis “systematics”. A *normalization* systematic just scales the overall rate of the process. A *Poisson* constraint adds a NP for each bin of the histogram.

A *shape systematic* defines “variation templates” to encode how much to penalize deviations from the expectation Gaussian constraint term in the likelihood. A single nuisance parameter correlates this post-fit prediction across the bins, as visualized in Figure 8.4. The pre-fit templates are shown in the pink and blue histograms (the choice of what defines the $\pm 1\sigma$ is arbitrary). The data are visualized in the black points, and the best fit NP to this data is shown in the grey histogram, which corresponds to a variation of -0.5σ .

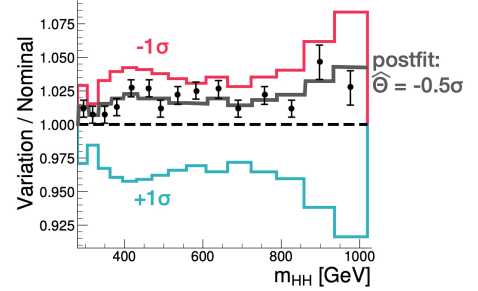


Figure 8.4: Gaussian constraint graphic (description in text).

We test if the fitted NPs are consistent with our constraints with the *pull*:

$$\text{pull} = \frac{\hat{\theta} - \theta_0}{\sigma_\theta}, \quad (8.6)$$

where the $\hat{\theta}$ are the post-fit NPs (which could also be $\hat{\hat{\theta}}(0)$ for the maximum likelihood fit with PoI fixed to 0), and θ_0 is the expected value from the constraint, and σ_θ is the error on the NP from the curvature of the likelihood.

The nuisance parameters are sometimes factorized as $\theta = [\theta_s, \theta_b]$, where θ_s are the nuisance parameters that only impact the signal model s , and θ_b only impact the background model b . The *background only fit* is the likelihood fit to the data with $\mu = 0$, so only the background nuisance parameters influence the fit.

9

HH Physics overview

The Higgs boson is a lynchpin of essentially the entire Standard Model.

– Dick Teresi, *The God Particle* (documentary of the Higgs discovery)

As introduced in Chapter 2, the terms in the SM related to EWK symmetry breaking are:

$$\mathcal{L}_h = |D_\mu \phi|^2 - V(\phi), \quad V(\phi) = -\mu^2 |\phi|^2 + \lambda |\phi|^4 \quad (9.1)$$

When ϕ is expanded for perturbations $h(x)$ about the vacuum expectation value (vev or v), the potential becomes:

$$V(\phi) \rightarrow \frac{1}{2} m_H^2 h(x)^2 + \lambda v h(x)^3 + \frac{1}{4} \lambda h(x)^2 \quad (9.2)$$

Measuring the Higgs mass and vev tells us everything about the SM Higgs sector, with $\lambda_{SM} = \sqrt{\frac{m_H^2}{2v^2}} = 0.129$.

This $V(\phi)$ is the simplest solution to give the vector bosons masses, but if nature does not have this simplest solution, then λ could deviate from this value. This is illustrated in Figure 9.1, where the blue line is the SM $V(\phi)$ and the yellow dot shows the broken symmetry state. The green and pink lines show expansions of $V(\phi)$ about the vev for λ differing from the SM expectation. Measuring λ independently of the Higgs mass allows us to test if the local curvature of the potential agrees with the SM.

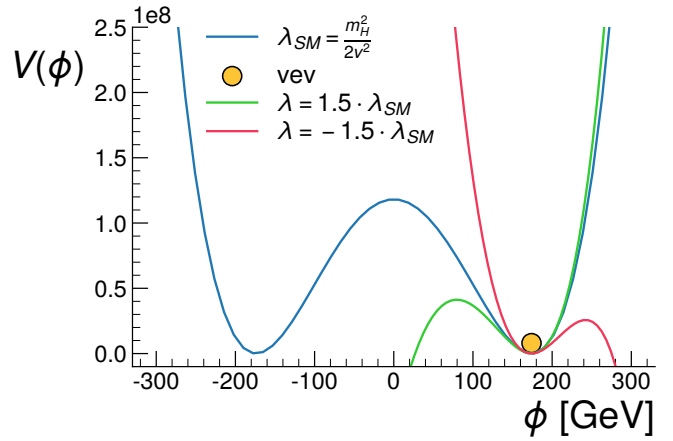


Figure 9.1: Impact of λ on the electroweak potential.

We can directly probe this $\lambda v h(x)^3$ interaction term by looking for the events where one Higgs boson produces two other Higgs bosons, called “di-Higgs production” or HH . Table 9.1 compares the dominant single and di-Higgs production cross sections for pp collisions at $\sqrt{s} = 13$ TeV. In the SM, HH events are three orders of magnitude rarer than single Higgs events, therefore we won’t be sensitive to the SM HH until our dataset grows by an order of magnitude with the HL-LHC dataset [135, 136]. We could see HH sooner if the Higgs self-coupling λ deviates from the SM expectation.

Prod mode	$\sigma(pp \rightarrow H + X)$ [pb]	$\sigma(pp \rightarrow HH + X)$ [fb]
ggF	48.57	31.05
VBF	3.782	1.73
Higgs-strahlung	2.267	0.865
ttH(H)	0.5071	0.775

Table 9.1: Cross-sections for single H and H production at $\sqrt{s} = 13$ TeV. The single Higgs cross-sections are in pb while the HH cross-sections are in fb. [137, 138]

Figure 9.2 delineates the main production modes for HH decay. While the Higgs boson was discovered in the clean decay channels of $H \rightarrow ZZ^*$ and $H \rightarrow \gamma\gamma$ [10, 11], since the HH cross-section is so much lower, all of the major HH channels include at least one of the Higgs bosons decaying to a $b\bar{b}$ pair. The $HH \rightarrow 4b$ analysis takes advantage of the leading HH branching ratio with 36.9% of the total cross-section, and the highlighted cross-sections in Table 9.1 indicate the HH production modes considered for the $4b$ analysis.

$b\bar{b}$	WW^*	gg	$\tau^+\tau^-$	$c\bar{c}$	ZZ^*	$\gamma\gamma$	
33.9%	24.9%	9.5%	7.3%	3.4%	3.1%	0.3%	$b\bar{b}$
	4.6%	3.5%	2.7%	1.2%	1.1%	< 0.1%	WW^*
		0.7%	1.0%	0.5%	0.4%	< 0.1%	gg
			0.4%	0.4%	0.3%	< 0.1%	$\tau^+\tau^-$
				< 0.1%	0.2%	< 0.1%	$c\bar{c}$
					< 0.1%	< 0.1%	ZZ^*
						< 0.1%	$\gamma\gamma$

Figure 9.2: HH branching ratios.

This chapter is organized as follows. In Chapter 9.1 details the κ framework used for the non-SM signal interpretations, and Chapter 9.2 details the signal reweighting procedure used to express these coupling variations. Chapter 9.3 introduces the Effective Field Theory interpretations also considered.

9.1 κ framework

In the SM, the dominant gluon-gluon-fusion (ggF) cross-section is $\sigma_{ggF}^{SM} = 31.05 \pm 3\%$ (PDF+ α_s) $^{+6\%}_{-23\%}$ (Scale + m_{top}) fb for a next-next-to-leading order (NNLO) calculation in the infinite top mass limit [139]. Figure 9.3 shows the two leading order (LO) diagrams for ggF production. In the box diagram (Figure 9.3(a)), the top loop radiates two separate Higgs bosons. The triangle diagram (Figure 9.3(b)) is mediated by the Higgs self-coupling as one Higgs produces the two other Higgses.

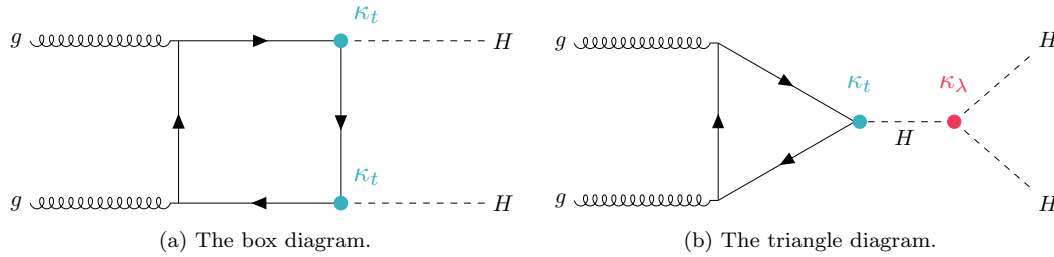


Figure 9.3: The leading order gluon-gluon fusion di-Higgs production Feynman diagrams.

The SM HH cross section is small partially because of a destructive interference between the box and triangle diagrams. Figure 9.4 show the differential cross section versus the invariant mass of the HH system for the box and triangle contributions as well as their interference. The contribution from the triangle diagram (in red) peaks at low m_{HH} . The box diagram has a harder spectrum with the rate increasing above $m_{HH} > 350\text{GeV} = 2m_t$ when the top-quarks in the loop are on-shell. The destructive interference between these two is shown in green and the lower differential cross section is indicated in the black line.

We search for non-SM λ couplings as deviations from the the SM value parametrized as $\kappa_\lambda = \lambda/\lambda_{SM}$. Increasing the $|\kappa_\lambda|$ increases the contribution of the triangle diagram, and can cause non-trivial interference shapes. Figure 9.5(a) illustrates how as $|\kappa_\lambda|$ becomes large, the total cross-section also increases as the triangle diagram dominates and its size increases. Figure 9.5 emphasizes how the interference terms translate into the m_{HH} shape. The SM contribution is shown in pink, but when $\kappa_\lambda = 2$ (in yellow), the destructive interference is close to maximal which induces a characteristic double-peak m_{HH} structure. For $\kappa_\lambda=10$ (in turquoise) the triangle diagram dominates and the m_{HH} spectrum becomes softer.

We also search for vector boson fusion (VBF) HH production, which has a SM cross-section

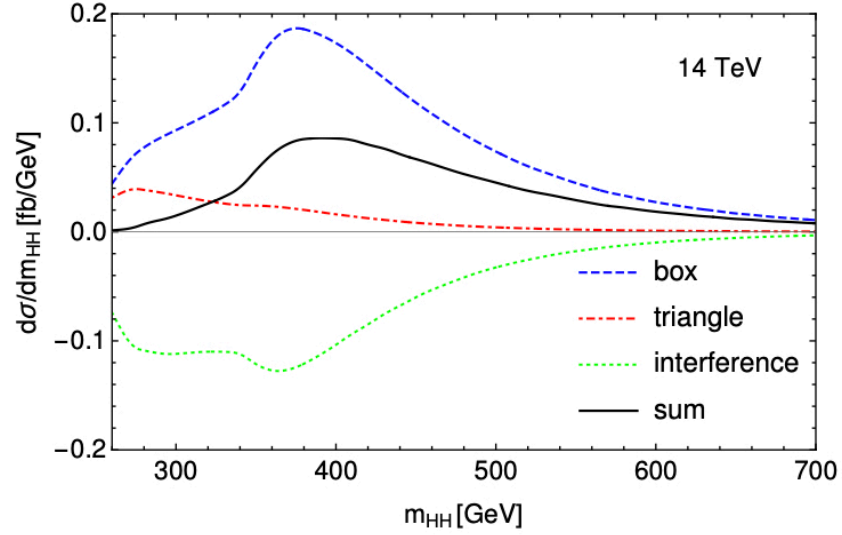
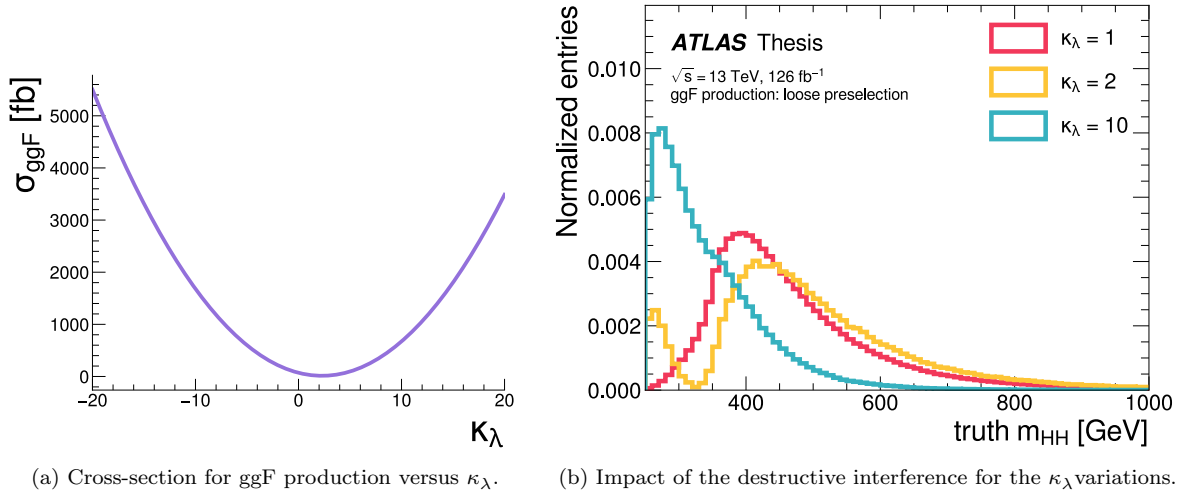


Figure 9.4: Impact of the interference of the box and triangle diagrams for ggF HH production, as a function of m_{HH} [138].



(a) Cross-section for ggF production versus κ_λ .

(b) Impact of the destructive interference for the κ_λ variations.

Figure 9.5: Impact of κ_λ variations on the ggF rate (left) and m_{HH} shape (right).

of $\sigma_{VBF HH}^{SM} = 1.726 \pm 2.1\%$ (PDF+ α_s) $^{+0.03\%}_{-0.04\%}$ (Scale) fb at N³LO [140]. This is over an order of magnitude smaller than the ggF HH cross-section. The LO VBF Feynman diagrams are shown in Figure 9.6, and we again interpret non-SM couplings in the κ framework, by parametrizing the deviations from the SM couplings. VBF production is also impacted by κ_λ (Figure 9.6(a)), but additionally gives us sensitivity to couplings of the vector bosons κ_{2V} and κ_V . Single Higgs analyses

already constrain κ_V , but HH production is the only place where we directly probe the κ_{2V} coupling.

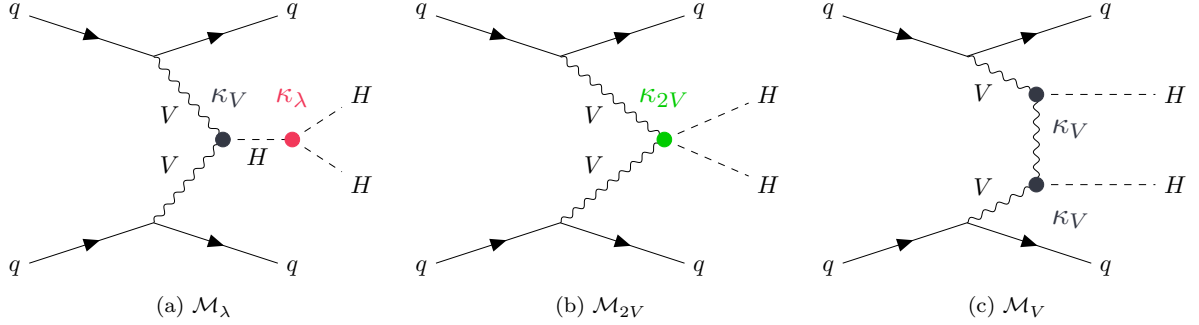


Figure 9.6: The three tree-level vector boson fusion di-Higgs production Feynman diagrams. A convention the matrix elements names is given in the captions of the respective diagrams.

Figure 9.7 shows the total cross section versus κ_λ (top) and κ_{2V} (bottom), again increasing with $|\kappa_\lambda|$ or $|\kappa_{2V}|$. The impact of these variations on these couplings on the VBF production m_{HH} shape is shown in Figure 9.8. The SM differential cross-section is shown in pink, and the κ_λ variations are shown in yellow ($\kappa_\lambda=2$) and turquoise ($\kappa_\lambda=10$). The κ_λ variation samples still have a softer m_{HH} spectrum. This is contrasted to the κ_{2V} variations in navy ($\kappa_{2V}=0$) and green ($\kappa_{2V}=1.5$), which have much heavier tails for high m_{HH} values.

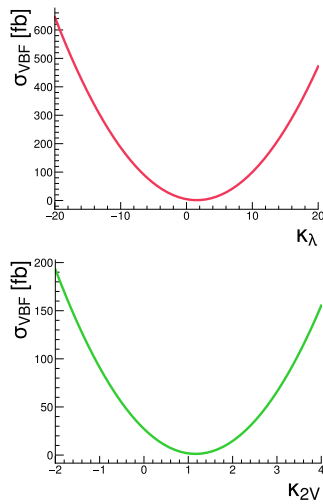


Figure 9.7: VBF production cross-sections versus κ_λ (top) and κ_{2V} (bottom).

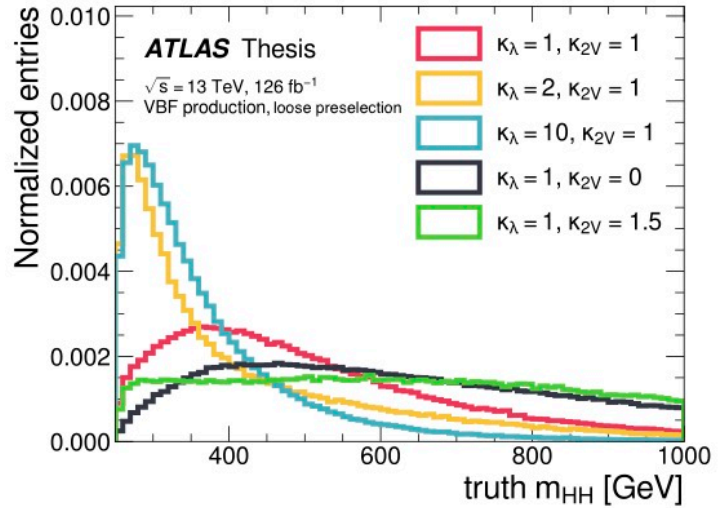


Figure 9.8: truth m_{HH} shape for selected VBF production signals with sm (pink), the κ_λ variations ($\kappa_\lambda=2$ in yellow and $\kappa_\lambda=10$ in turquoise), and κ_{2V} variations ($\kappa_{2V}=0$ in navy and $\kappa_{2V}=1.5$ in green).

9.2 Signal parametrization

Although it is possible to generate a signal sample for any κ_λ , it is computationally too expensive to simulate these samples for every signal point in a κ_λ scan. We can generate the differential cross-section for any process by exchanging the parametrization for the three terms (box, triangle, and interference) that characterize the cross-section for three “basis functions”. This is done by solving a set of three linear equations. We work through the math in Chapter 9.2.2 for the ggF case. To characterize the variations of the three VBF couplings, the same principle can be applied, but involves solving a set of six linear equations, and we just state the result for VBF production.

9.2.1 ggF: histogram based reweighting

Although Figure 9.3 shows the LO Feynman diagrams, there is a fully differential next-to-leading-order (NLO) calculation. The terms in the NLO calculation still break down into diagrams in the box and triangle families. The sum of the terms in the box and triangle diagrams will be referred to as B and T , respectively. The impact from non-SM couplings come as a multiplicative factors for the corresponding vertices, allowing us to write down a combined amplitude (parametrized by κ_t and κ_λ as:

$$\mathcal{A}(\kappa_t, \kappa_\lambda) = \kappa_t^2 B + \kappa_t \kappa_\lambda T. \quad (9.3)$$

The ggF cross section is obtained by squaring this amplitude:

$$\sigma(pp \rightarrow HH) = |\mathcal{A}(\kappa_t, \kappa_\lambda)|^2 = \left(\kappa_t^2 B^* + \kappa_t \kappa_\lambda T^* \right) \left(\kappa_t^2 B + \kappa_t \kappa_\lambda T \right) \quad (9.4)$$

$$= \kappa_t^4 \left[|B|^2 + \frac{\kappa_\lambda}{\kappa_t} (B^* T + T^* B) + \left(\frac{\kappa_\lambda}{\kappa_t} \right)^2 |T|^2 \right]. \quad (9.5)$$

The κ_t^4 term scales the rate of the process. The 2^{nd} order κ_λ/κ_t polynomial dictates the way these diagrams interfere, so with three different κ_λ values, we can get any arbitrary κ_λ value. Consider the basis functions with $\kappa_\lambda = 0$ (no triangle diagram), $\kappa_\lambda = 1$ (SM), and a final (now arbitrary) $\kappa_\lambda = \kappa_0$. We keep $\kappa_t = 1$ because HH searches are not as sensitive to the top Yukawa as ttH analyses.

Plugging these basis points into Eq. 9.5 we get three different differential cross-sections:

$$|\mathcal{A}(1, 0)|^2 = |B|^2 \quad (9.6)$$

$$|\mathcal{A}(1, 1)|^2 = |B|^2 + (B^* T + T^* B) + |T|^2 \quad (9.7)$$

$$|\mathcal{A}(1, \kappa_0)|^2 = |B|^2 + \kappa_0 (B^* T + T^* B) + \kappa_0^2 |T|^2 \quad (9.8)$$

Now we solve for $|B|^2$, $(B^* T + T^* B)$, and $|T|^2$ as a function of $|\mathcal{A}(1, 0)|^2$, $|\mathcal{A}(1, 1)|^2$, and $|\mathcal{A}(1, \kappa_0)|^2$.

Define

$$\begin{cases} x = |B|^2 \\ y = (B^*T + T^*B) \\ z = |T|^2 \end{cases} \quad \text{and} \quad \begin{cases} a = |\mathcal{A}(1,0)|^2 \\ b = |\mathcal{A}(1,1)|^2 \\ c = |\mathcal{A}(1,\kappa_0)|^2 \end{cases} \quad (9.9)$$

and write this as a system of linear equations:

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & \kappa_0 & \kappa_0^2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad \Rightarrow \quad a = x \quad \text{and} \quad \begin{bmatrix} b - a \\ c - a \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \kappa_0 & \kappa_0^2 \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} \quad (9.10)$$

Then we take the inverse of this matrix to solve for the remaining two terms.

$$\begin{bmatrix} y \\ z \end{bmatrix} = \frac{1}{\kappa_0(\kappa_0 - 1)} \begin{bmatrix} \kappa_0^2 & -1 \\ -\kappa_0 & 1 \end{bmatrix} \begin{bmatrix} b - a \\ c - a \end{bmatrix} \quad (9.11)$$

$$y = \frac{1}{\kappa_0(\kappa_0 - 1)} [\kappa_0^2(b - a) - (c - a)] = -\frac{\kappa_0 + 1}{\kappa_0}a + \frac{\kappa_0^2}{\kappa_0(\kappa_0 - 1)}b - \frac{1}{\kappa_0(\kappa_0 - 1)}c \quad (9.12)$$

$$z = \frac{1}{\kappa_0(\kappa_0 - 1)} [-\kappa_0(b - a) + (c - a)] = \frac{1}{\kappa_0}a - \frac{1}{\kappa_0 - 1}b + \frac{1}{\kappa_0(\kappa_0 - 1)}c \quad (9.13)$$

Now we are ready to plug these expressions for the box, triangle, and interference terms (x , y , and z) to solve for $|\mathcal{A}(\kappa_t, \kappa_\lambda)|^2$ in terms of Eq. 9.5 in terms of three known cross-sections (a , b , and c):

$$\begin{aligned} |\mathcal{A}(\kappa_t, \kappa_\lambda)|^2 &= \kappa_t^2 \left[\kappa_t^2 |\mathcal{A}(1,0)|^2 \right. \\ &\quad + \kappa_t \kappa_\lambda \left(-\frac{\kappa_0 + 1}{\kappa_0} |\mathcal{A}(1,0)|^2 + \frac{\kappa_0^2}{\kappa_0(\kappa_0 - 1)} |\mathcal{A}(1,1)|^2 - \frac{1}{\kappa_0(\kappa_0 - 1)} |\mathcal{A}(1,\kappa_0)|^2 \right) \\ &\quad \left. + \kappa_\lambda^2 \left(\frac{1}{\kappa_0} |\mathcal{A}(1,0)|^2 - \frac{1}{\kappa_0 - 1} |\mathcal{A}(1,1)|^2 + \frac{1}{\kappa_0(\kappa_0 - 1)} |\mathcal{A}(1,\kappa_0)|^2 \right) \right] \end{aligned} \quad (9.14)$$

We reshuffle to put the coefficients in front of the basis functions together:

$$|\mathcal{A}(\kappa_t, \kappa_\lambda)|^2 = \kappa_t^2 \left[\left(\kappa_t^2 + \frac{\kappa_\lambda^2}{\kappa_0} - \frac{1 + \kappa_0}{\kappa_0} \kappa_t \kappa_\lambda \right) |\mathcal{A}(1,0)|^2 + \frac{\kappa_0 \kappa_t \kappa_\lambda - \kappa_\lambda^2}{\kappa_0 - 1} |\mathcal{A}(1,1)|^2 + \frac{\kappa_\lambda^2 - \kappa_\lambda \kappa_t}{\kappa_0(\kappa_0 - 1)} |\mathcal{A}(1,\kappa_0)|^2 \right]. \quad (9.15)$$

The ATLAS HH analyses use the above prescription with $\kappa_0 = 20$ as the last basis function. So simplifying Eq. 9.5 for the combination formula for our implementation:

$$|\mathcal{A}(\kappa_\lambda)|^2 = \left[\left(1 - \frac{21}{20}\kappa_\lambda + \frac{1}{20}\kappa_\lambda^2\right) \underline{|\mathcal{A}(1,0)|^2} + \frac{\kappa_\lambda(20 - \kappa_\lambda)}{19} \underline{|\mathcal{A}(1,1)|^2} + \frac{\kappa_\lambda(\kappa_\lambda - 1)}{380} \underline{|\mathcal{A}(1,\kappa_0)|^2} \right]. \quad (9.16)$$

Although Eq. 9.16 is fully differential, we just use the truth m_{HH} differential distribution to derive reweighting functions mapping from the SM ($\kappa_\lambda = 1$) to each other κ_λ signal we test. These signal reweighting functions are derived at the parton level, before the hadronization, and interaction with the detector effects. These κ_λ reweighting functions were derived with over a million truth events simulated with Powheg Box v2, in the full phase space. This is a simplification to just take into account the m_{HH} performance of the fully differential cross-section, but it is an advantageous one because it means we don't need to simulate three large statistics basis samples with the full detector simulation analysis chain, and instead can just use the SM sample (which had 3.8 million simulated events) to get the rest of the κ_λ points. To check the veracity of this assumption and the validity in the phase space of the analysis, a $\kappa_\lambda = 10$ sample (with 1.9 million simulated events) was produced and passed through the whole simulation chain, and the modeling of the reconstruction level variables was checked. Figure 9.9 shows the signal reweighting performance for the reconstructed m_{HH} (left) and pseudo-rapidity between the Higgses (right) in the 4b signal region (i.e, after the analysis cuts which will be described in Chapter 10). The truth m_{HH} reweighting is shown in turquoise, and matches well with the simulated $\kappa_\lambda = 10$ prediction shown in navy. Also, the limits for $\kappa_\lambda = 10$ sample versus the reweighted $\kappa_\lambda = 10$ sample were consistent with each other at the %-level, which justified this choice of expressing all of the kinematics with the truth m_{HH} parametrization.

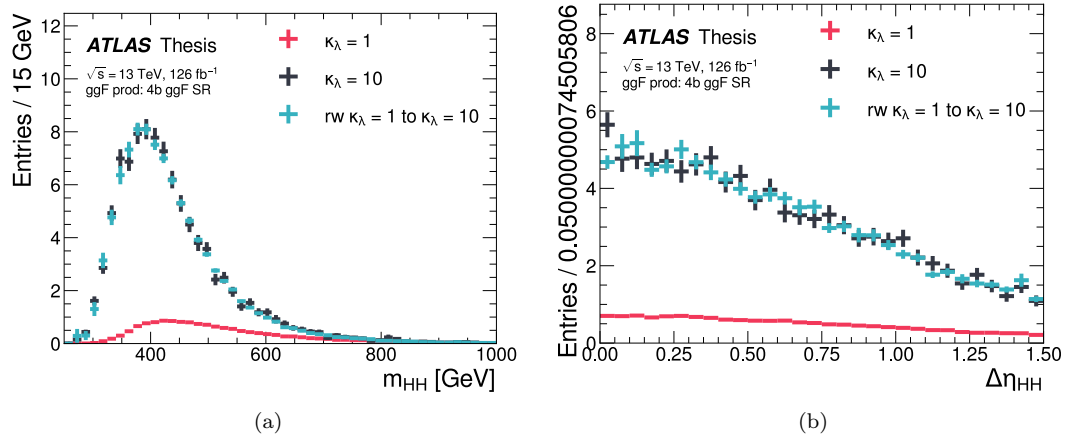


Figure 9.9: Impact of the signal reweighting for the $\kappa_\lambda=10$ sample for the reconstructed m_{HH} (left) and pseudo-rapidity between the reconstructed Higgs bosons (right). The SM that we reweighted from is shown in pink, while the reweighted distribution for $\kappa_\lambda=10$ is shown in turquoise and compared to true $\kappa_\lambda=10$ distribution in navy. The error bars are statistical only.

Figure 9.9 also shows the SM distribution (in pink) which is the source distribution for the reweighting. The statistical error on the reweighted templates gets larger in regions of low support for the SM distribution, as seen by the larger turquoise error bars for the lowest m_{HH} values. For the ggF signal, these reweighting errors didn't impact the analysis, but this choice of basis functions to avoid unphysical signal templates was a more challenging task for the VBF analysis.

9.2.2 VBF: event level reweighting

For the VBF differential cross-section at arbitrary couplings, the methodology is extended for the case of three diagrams. These are shown in Figure 9.6, where $|\mathcal{M}_\lambda|$ is the diagram that depends on κ_λ , $|\mathcal{M}_{2V}|$ is the diagram that depends on κ_{2V} , and $|\mathcal{M}_V|$ is the diagram with two Higgses radiations off of the vector boson mediated by κ_V .

Writing out the differential cross-section and taking into account the scalings with respect to the couplings gives:

$$\begin{aligned} \sigma(\kappa_\lambda, \kappa_{2V}, \kappa_V) &= \left| \kappa_V \kappa_\lambda \mathcal{M}_\lambda + \kappa_{2V} \mathcal{M}_{2V} + \kappa_V^2 \mathcal{M}_V \right|^2 \\ &= \kappa_V^2 \kappa_\lambda^2 |\mathcal{M}_\lambda|^2 + \kappa_V \kappa_\lambda \kappa_{2V} (\mathcal{M}_\lambda \mathcal{M}_{2V}^* + \mathcal{M}_\lambda^* \mathcal{M}_{2V}) + \kappa_V^3 \kappa_\lambda (\mathcal{M}_\lambda \mathcal{M}_V^* + \mathcal{M}_\lambda^* \mathcal{M}_V) \\ &\quad + \kappa_{2V}^2 |\mathcal{M}_{2V}|^2 + \kappa_{2V} \kappa_V^2 (\mathcal{M}_{2V} \mathcal{M}_V^* + \mathcal{M}_{2V}^* \mathcal{M}_V) + \kappa_V^2 |\mathcal{M}_V|^2. \end{aligned} \quad (9.17)$$

With the six terms, we need six linearly independent $(\kappa_\lambda, \kappa_{2V}, \kappa_V)$ couplings for basis functions to express the kinematics. Although with infinite statistics we would be free to choose any 6 linearly independent couplings to find a basis, with finite statistics, it's important to choose a set of basis functions that is well representative of the BSM phase space to avoid non-physical signal templates. The chosen basis samples are shown in Table 9.2.

κ_λ	κ_{2V}	κ_V
1	1	1
1	1.5	1
2	1	1
10	1	1
1	1	0.5
-5	1	0.5

Table 9.2: Coupling values defining the basis functions for the VBF signal reweighting.

Solving the system of linear equations with these couplings as bases gives the differential cross-section

for arbitrary couplings:

$$\begin{aligned}
\sigma(\kappa_\lambda, \kappa_{2V}, \kappa_V) = & \left(\frac{68}{135} \kappa_{2V}^2 - 4 \kappa_{2V} \kappa_V^2 + \frac{20}{27} \kappa_{2V} \kappa_V \kappa_\lambda + \frac{772}{135} \kappa_V^4 - \frac{56}{27} \kappa_V^3 \kappa_\lambda + \frac{1}{9} \kappa_V^2 \kappa_\lambda^2 \right) \sigma(1, 1, 1) \\
& + \left(-\frac{4}{5} \kappa_{2V}^2 + 4 \kappa_{2V} \kappa_V^2 - \frac{16}{5} \kappa_V^4 \right) \sigma(1, 1.5, 1) \\
& + \left(\frac{11}{60} \kappa_{2V}^2 + \frac{1}{3} \kappa_{2V} \kappa_V^2 - \frac{19}{24} \kappa_{2V} \kappa_V \kappa_\lambda - \frac{53}{30} \kappa_V^4 + \frac{13}{6} \kappa_V^3 \kappa_\lambda - \frac{1}{8} \kappa_V^2 \kappa_\lambda^2 \right) \sigma(2, 1, 1) \\
& + \left(-\frac{11}{540} \kappa_{2V}^2 + \frac{11}{216} \kappa_{2V} \kappa_V \kappa_\lambda + \frac{13}{270} \kappa_V^4 - \frac{5}{54} \kappa_V^3 \kappa_\lambda + \frac{1}{72} \kappa_V^2 \kappa_\lambda^2 \right) \sigma(10, 1, 1) \\
& + \left(\frac{88}{45} \kappa_{2V}^2 - \frac{16}{3} \kappa_{2V} \kappa_V^2 + \frac{4}{9} \kappa_{2V} \kappa_V \kappa_\lambda + \frac{152}{45} \kappa_V^4 - \frac{4}{9} \kappa_V^3 \kappa_\lambda \right) \sigma(1, 1, 0.5) \\
& + \left(\frac{8}{45} \kappa_{2V}^2 - \frac{4}{9} \kappa_{2V} \kappa_V \kappa_\lambda - \frac{8}{45} \kappa_V^4 + \frac{4}{9} \kappa_V^3 \kappa_\lambda \right) \sigma(-5, 1, 0.5)
\end{aligned}$$

Although the ggF signal reweighting had good closure for a one-dimensional truth m_{HH} parametrization, for the VBF analysis m_{HH} didn't capture the full kinematics of the event with the analysis cuts and final discriminant. Thus, the VBF signal reweighting is done at the event level using this linear combination of signal samples.

9.3 EFT interpretations

In addition to the κ -framework for parametrizing the new physics, we also considered more generic signals predicted by Effective Field Theory (EFT) frameworks. If the new physics is at an energy scale Λ large enough that the new particles can't be observed directly, the additional diagrams can still give residual corrections at lower energies which will give deviations from the SM expectation. The Effective Field Theory framework is a way to systematically parametrize the new physics that could give rise to such lower energy deviations. We set constraints in both the SMEFT and HEFT frameworks, which are briefly outlined below.

9.3.1 SMEFT

The Standard Model EFT (SMEFT) organizes the extra terms that can be added in terms of their mass dimensions [141]. For the Lagrangian density to stay dimension 4, the extra higher dimensional operators come accompanied by powers of $1/\Lambda^{d-4}$. The dimension 5 operator violates lepton number conservation, so we begin by considering operators with mass dimension six giving the first SMEFT extension of the Lagrangian as:

$$\mathcal{L}_{SMEFT} = \mathcal{L}_{SM} + \frac{1}{\Lambda^2} \sum_i c_i \mathcal{O}_i^{(6)} + \dots \quad (9.18)$$

where the c_i are the Wilson coefficients, or couplings naturally of order 1, and the $\mathcal{O}_i^{(6)}$ are the dimension six operators. For analyses probing the m_H scale, variations due to $\Lambda = 1$ TeV scale new physics will lead to $\mathcal{O}(10\%)$ effects in analysis observables. 10% is already lower than the present HH sensitivity to coupling variations, which justifies only using dimension six operators for the HH interpretations.

In SMEFT, the Higgs field is still an $SU(2)$ doublet, and the full SMEFT Lagrangian is invariant under the $SU(2) \times U(1)$ transformations. When assuming the deviations only impact a single fermion generation, there are 59 independent dimension-six operators. For HH analyses, we only consider the five (CP-conserving) terms that directly impact the HH cross-section. We're considering the SMEFT parameters in the Warsaw basis [142], with the EFT couplings characterized by: c_H , c_{tH} , c_{tG} , c_{HG} . The dimension six piece of the SMEFT Lagrangian is given in Eq. 9.19,

$$\begin{aligned} \Delta\mathcal{L}_{SMEFT}^{(6)} = & \frac{c_{H\Box}}{\Lambda^2} \partial_\mu(\phi^\dagger\phi)\partial^\mu(\phi^\dagger\phi) + \frac{c_H}{\Lambda^2} (\phi^\dagger\phi)^3 + \frac{c_{tH}}{\Lambda^2} (\phi^\dagger\phi\bar{Q}_L\tilde{\phi}t_R + \text{h.c.}) \\ & + \frac{c_{HG}}{\Lambda^2} \phi^\dagger\phi G_{\mu\nu}^a G^{a,\mu\nu} + \frac{c_{tG}}{\Lambda^2} ig_s(\bar{Q}_L\sigma_{\mu\nu}T^a\tilde{\phi}t_R G_a^{\mu\nu}) \end{aligned} \quad (9.19)$$

where “h.c.” is the Hermitian conjugate, and $\tilde{\phi} \equiv i\sigma_2\phi$. c_H directly shifts the HHH vertex, while $c_{H\Box}$ rescales all of the Higgs couplings. These two can be related to the shift caused in κ_λ as:

$$\kappa_\lambda = 1 - 2\frac{v^4}{m_H^2\Lambda^2}c_H - 3\frac{v^2}{\Lambda^2}c_{H\Box} \quad (9.20)$$

Figure 9.10 shows the Feynman diagrams for the new couplings and modifications.

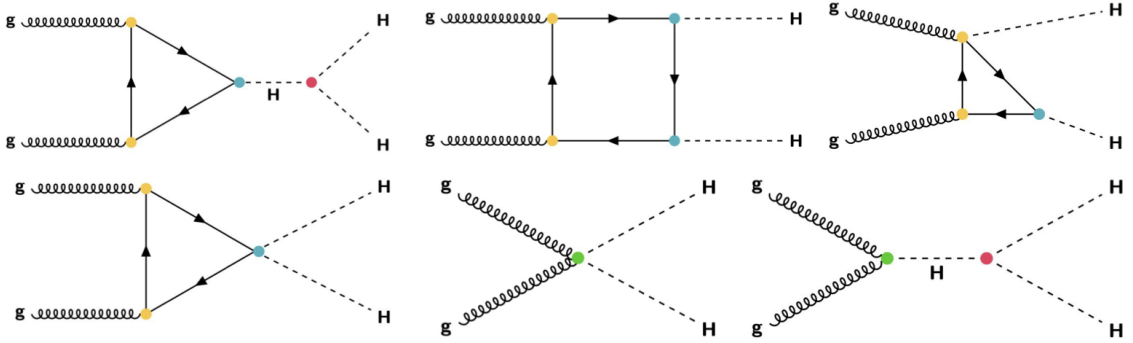


Figure 9.10: Illustration of the coupling modifications in the SMEFT framework [143].

9.3.2 HEFT

Another way to parametrize the new physics is using the Higgs Effective Field Theory (HEFT)[144, 145]. The electroweak sector is now a non-linear realization of a $U(1) \times SU(2) \times SU(2)$ symmetry.

The mechanism of electroweak symmetry breaking is now guided by chiral perturbation theory, and the Higgs field is now a singlet of the unbroken $U(1) \times SU(2)$ symmetry. The anomalous couplings factorize to allow for a simplified treatment for the HH contributions in a 5d parameter space with Wilson coefficients: c_{hhh} , c_{tth} , c_{ggh} , c_{gghh} , and c_{tthh} . The c_{hhh} and c_{tth} , correspond to the dimension 4 operators in Figure 9.3, with $c_{hhh} = \kappa_\lambda$ and $c_{tth} = \kappa_t$. There are additionally other diagrams corresponding to dimension 6 operators that don't exist in the SM, shown in Figure 9.11. The extended Lagrangian corresponding to these HEFT couplings is:

$$\Delta\mathcal{L}_{\text{HEFT}} = -m_t \left(c_{tth} \frac{h}{v} + c_{tthh} \frac{h^2}{v^2} \right) \bar{t}t - c_{hhh} \frac{m_h^2}{2v} h^3 + \frac{\alpha_s}{8\pi} \left(c_{ggh} \frac{h}{v} + c_{gghh} \frac{h^2}{v^2} \right) G_{\mu\nu}^a G^{a,\mu\nu}. \quad (9.21)$$

In the SM, $c_{hhh} = c_{tth} = 1$ and $c_{ggh} = c_{gghh} = c_{tthh} = 0$.

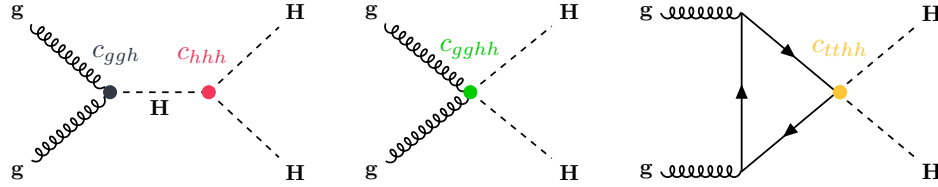


Figure 9.11: Illustration of the coupling modifications for the BSM (dimension 6) diagrams in the HEFT framework [146].

HEFT is a strictly more general framework than SMEFT, which can make it more challenging to disentangle which operator might be generating the new physics (i.e, distinguishing whether an enhancement in the HH production rate is coming from c_{tthh} or c_{hhh}). In practice, to probe the 5 dimensional parameter space, we use the set of benchmarks from the cluster analysis [147]. The cluster analysis for the benchmarks in Table 9.3 uses an NLO HEFT calculation with the full m_t mass dependence included [148]. Coefficients of the 5d Wilson coefficients are sampled from the allowed ranges, and then used to search for which of the coupling choices have similar m_{HH} shapes. A “benchmark model” is chosen to define the representative shape in each cluster, and the coupling choices for the seven benchmark models are shown in Table 9.3. The m_{HH} shapes for each of these benchmarks are visualized in Figure 9.12.

EFT signal parametrization

Both the SMEFT and HEFT signal variations are characterized by the truth m_{HH} based reweighting derived at the parton level. Only the ggF signal is considered in the EFT interpretations.

Benchmark	c_{hhh}	c_{tth}	c_{ggh}	c_{gggh}	c_{tthh}
SM	1	1	0	0	0
BM 1	3.94	0.94	1/2	1/3	-1/3
BM 2	6.84	0.61	0	-1/3	1/3
BM 3	2.21	1.05	1/2	1/2	-1/3
BM 4	2.79	0.61	-1/2	1/6	1/3
BM 5	3.95	1.17	1/6	-1/2	-1/3
BM 6	5.68	0.83	-1/2	1/3	1/3
BM 7	-0.10	0.94	1/6	-1/6	1

Table 9.3: The coefficient values corresponding to the BSM templates, as proposed by [148]

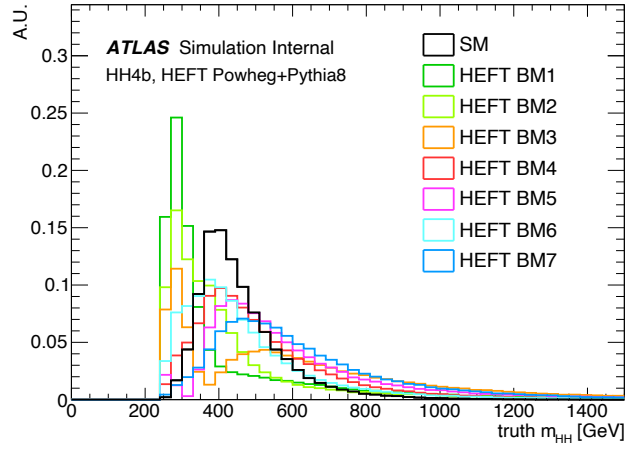


Figure 9.12: The SM and HEFT benchmark signals. The benchmarks correspond to the coupling values shown in Table 9.3.

10

Analysis selection

In total, inclusive diHiggs production with decay to four b quarks has a signal-over-background ratio S/B which is too bad to be a suitable search channel. . .

– Matthew Dolan, Christoph Englert, and Michael Spannowsky (2012) [149]

While the above quote sets the stage for the challenging prospects of this final state – the last decade of work on both ATLAS and CMS has firmly established $4b$ as a strongly contributing channel for the HH combination. Figure 10.1 illustrates the analysis steps and the variables will be discussed and described in this chapter. Chapter 10.1 discusses the triggers, while Chapter 10.2 describes the jet selection (Steps 2–3). Chapter 10.3 summarizes the event reconstruction with the the VBF / ggF channel orthogonalization, Higgs Candidate reconstruction, and background rejection cuts.

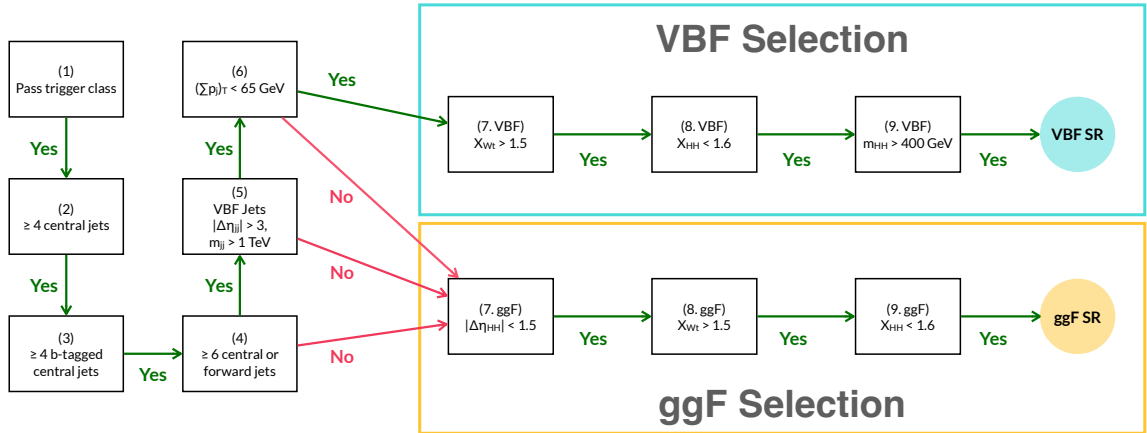


Figure 10.1: Flow chart for the $HH \rightarrow 4b$ analysis. The first three steps need to be passed for an event to be in either analysis category. Events passing steps (4) – (6), go in the “VBF Selection” channel, while any events failing these cuts go in the “ggF Selection” channel. (7) – (9) are the final cuts that define the VBF and ggF SR.

Chapter 10.4 summarizes the cutflows of this workflow, and the discriminant categorization is discussed in Chapter 10.5.

10.1 Triggers

This analysis uses a combination of multi b -jet triggers. The p_T thresholds and b -tagging working points vary slightly by the year of data taking (with the specific cut values delineated in Table 10.1). Only 2 b -tags are required in the trigger to avoid creating a bias in the $2b$ the control region used to define the background estimate (to be described in Chapter 11.1. The b -tagging SFs are derived for each trigger chain individually, so we define “trigger categories” to apply the trigger SFs, described in Chapter 10.1.1. Also, the trigger are not fully efficient with respect to the offline analysis cuts, illustrated as the efficiencies are less than 100% in Figure 10.2, and also this efficiency varies as a function of the reconstructed 4-jet invariant mass. To account for this feature of “operating on the turn on curve” the SF accounts for the trigger effects from the online b tagging (Chapter 10.1.2) and the E_T kinematic differences (Chapter 10.1.3).

Trigger Type	Year	HLT thresholds	L1 thresholds
2b1j	2016	$p_T > 100$ GeV jet & two $p_T > 55$ GeV 60% WP b -jets	five $p_T > 15$ GeV jets
	2017	one $p_T > 150$ GeV jet and two $p_T > 55$ GeV 70% WP b -jets	one $p_T > 85$ GeV jet and two $p_T > 30$ GeV jets
	2018	one $p_T > 150$ GeV jet and two $p_T > 55$ GeV 70% WP b -jets	one $p_T > 85$ GeV jet and two $p_T > 30$ GeV jets
2b2j	2016	four $p_T > 35$ GeV jets, two 60% WP b -tags	four $p_T > 15$ GeV, $ \eta < 2.5$ jets
	2017	four $p_T > 35$ GeV jets, two 40% WP b -tags	
	2018	four $p_T > 35$ GeV jets, two 60% WP b -tags	

Table 10.1: Triggers used for non-resonant searches. For b -tagging in the trigger in Run 2, the MV2 version of the b -tagger is used. Also, an L1 $|\eta| < 3.2$ cut is assumed where not specified.

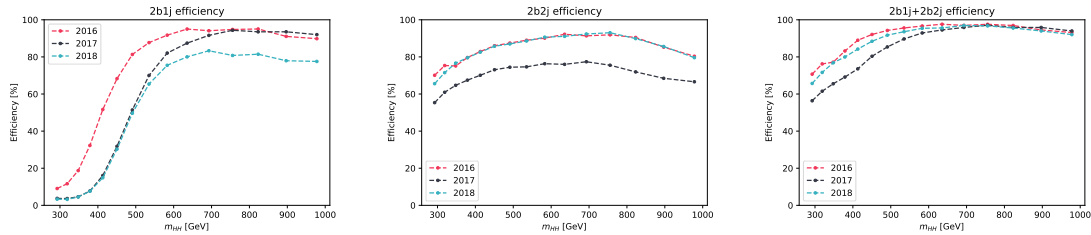


Figure 10.2: Trigger efficiencies of the 2b1j, 2b2j and combined for the MC16a/d/e corresponding to years 2016-2018 for the SM ggF $\kappa_\lambda=1$ signal. Significantly lower efficiency for 2017 2b2j compared to other years is due to tighter b -tagging requirement (40% for 2017, 60% for 2016 and 2018).

10.1.1 Trigger buckets

Since the trigger chains are built up from the p_T and b -tagging cuts on the jets, the event level SFs is the product over the jet level SFs, as shown in Eq. 10.1: with two contributions:

$$\text{Multi } b\text{-jet trigger SF} = \prod_i^N SF_{jet}^{b\text{-tag}}(i) \times \prod_j^4 SF_{jet}^{kinematic}(i), \quad \text{with} \quad (10.1)$$

- b -jet trigger SFs (described in Chapter 10.1.2)
- Kinematic E_T HLT and L1 SFs (described in Chapter 10.1.3)

To reconstruct the trigger decision and define the jet level SFs, offline jets are matched to the online jets using a ΔR matching criterion. These online jets are then checked to pass the (online) thresholds given in Table 10.1, and if this many jets and b -jets pass this selection, the event passes this trigger.

The online jet b -tagging WPs can vary by the trigger type, for example, the 2018 triggers use the 70% b -tagging WP for the 2b1j trigger, but the 60% WP for the 2b2j trigger. To decide whether to use the 70% or 60% online b -tag SF, “trigger buckets” are created which separate the events based on the offline jet p_T s to decide which trigger chain to check. The cuts defining the buckets are designed to mimic the trigger kinematics to retain maximal efficiency. Trigger **bucket 1** mimics the 2b1j trigger by cutting on the offline jets $p_{T,1} > 170$ GeV and $p_{T,3} > 70$ GeV, with the jets are ordered by p_T . An event that fails these jet cuts goes in trigger **bucket 2**. In trigger bucket 1, we keep the event if the 2b1j trigger passes, and in trigger bucket 2, we check the 2b2j trigger. Then based on the trigger chain that was checked the corresponding b -tagging SF is applied. Figure 10.3 summarizes this procedure.

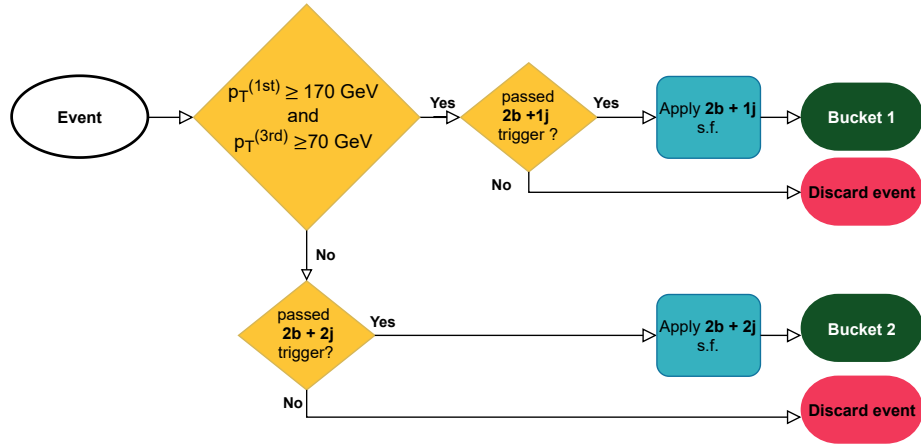


Figure 10.3: Trigger bucket strategy for the $HH \rightarrow 4b$ non-resonant search.

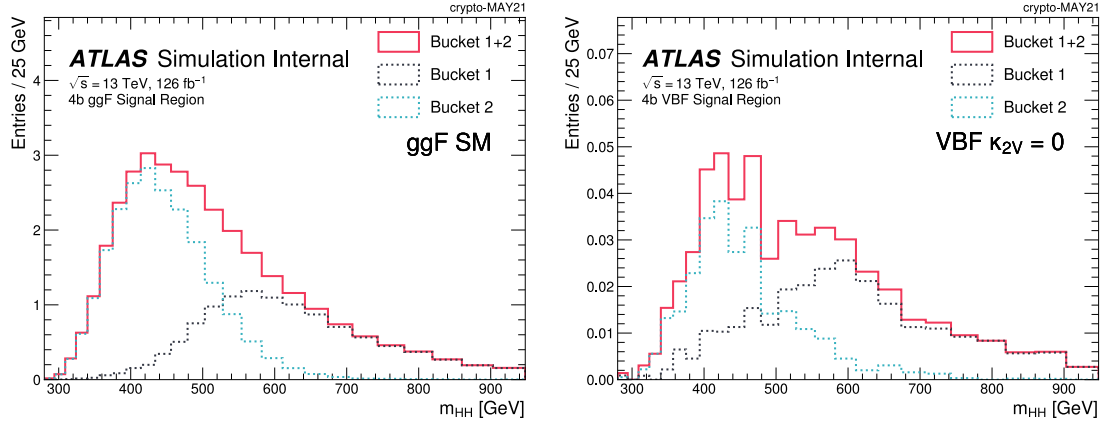


Figure 10.4: The bucket composition of m_{HH} for the SM ggF (left) and $\kappa_{2V}=0$ VBF (right) HH MC simulation in the 4b Signal Regions. Bucket 1 corresponds to the 2b1j trigger and Bucket 2 corresponds to the 2b2j trigger.

Figure 10.4 shows how this strategy of using a combination of two triggers gives us sensitivity to complementary phase spaces in the analysis. The 2b1j trigger drives our acceptance for the high m_{HH} events, while the 2b2j trigger provides our low m_{HH} acceptance.

10.1.2 b -jet SF

The offline and online b -tagging decisions are highly correlated, so the online b -tagging SF are derived conditionally based on the offline b -tagging decision [91]. The online b -tagging SFs derivation follows the recipe of the offline b -jet calibration Chapter 6.5 with a fully leptonic $t\bar{t}$ decay with opposite sign electrons and muons to increase the purity. Denoting passing the online working point as “on” and passing the offline efficiency as “off”, the efficiency for passing the online b -tag WP given the offline working point is passed is: $\varepsilon(\text{on}|\text{off})$. The online or offline WP could also fail, and these cases are denoted as: $\overline{\text{on}}$ and $\overline{\text{off}}$. The combination of offline and online b -tagging decisions gives four cases:

- Case 1: Pass online and offline b -tagging:

$$\varepsilon(\text{on} \wedge \text{off}) = \varepsilon(\text{on}|\text{off})\varepsilon(\text{off})$$
- Case 2: Fail the online b -tag, but pass the offline b -tag:

$$\varepsilon(\overline{\text{on}} \wedge \text{off}) = [1 - \varepsilon(\text{on}|\text{off})]\varepsilon(\text{off})$$
- Case 3: Pass the online b -tag, but fail the offline b -tag:

$$\varepsilon(\text{on} \wedge \overline{\text{off}}) = \varepsilon(\text{on}) - \varepsilon(\text{on}|\text{off})\varepsilon(\text{off})$$
- Case 4: Fail the online and offline b -tagging:

$$\varepsilon(\overline{\text{on}} \wedge \overline{\text{off}}) = 1 - \varepsilon(\text{off}) - \varepsilon(\text{on}) + \varepsilon(\text{on}|\text{off})\varepsilon(\text{off})$$

Then for each efficiency, we still apply $SF = \varepsilon^{data}/\varepsilon^{MC}$. For offline jets that are not matched to a corresponding online HLT jet, just the offline SF is applied, just the offline b -tagging SF is applied.

The data quality of b -jet triggers dictates SFs dictates the Run 2 dataset usable in this analysis.

1. In 2016 there was an issue in the online beam spot calculation, which impacted the primary vertex calculation for the HLT b -tagging. This results in a loss of 8.3 fb^{-1} from the 32.8 fb^{-1} of the full 2016 dataset.
2. For 2017 and 2018, the first luminosity blocks of data taking where the beam spot has yet updated are also discarded. This gives $\approx 1.5\%$ lower luminosity in these years than the baseline luminosity [91].
3. The 2015 triggers are not included in Table 10.1 as the 2015 conditional b -jet trigger SFs were not available for DL1r b -tagger.

When accounting for the above three points, Table 10.2 is the luminosity for each year of the 4b analysis, with a total luminosity is 126.0 fb^{-1} .

Year	Luminosity [fb^{-1}]
2016	24.6
2017	43.7
2018	57.7
all	126.0

Table 10.2: Luminosity by year for the $HH \rightarrow 4b$ analysis.

10.1.3 Kinematic SF

To account for the analysis operating on the turn-on curve, kinematic SFs are derived to account for the differences in the L1 and HLT efficiencies in data and MC. The HLT (or L1) trigger can be described through a sequence of decisions for each of the jets passing the corresponding p_T thresholds (delineated in Table 10.1). For example, the 2017 2b1j trigger decision can be decomposed into three sub-decisions for each jet passing a trigger threshold: the leading jet passing $p_T > 100 \text{ GeV}$, and the 2^{nd} and 3^{rd} leading jets pass $p_T > 55 \text{ GeV}$. The jet p_T SFs are derived separately and then the event level kinematic SF is the product over the jet p_T SFs. For each jet, both an HLT and L1 SF is derived.

These SFs were derived in muonic $t\bar{t}$ decays, with the following selection:

- A single muon trigger is used with exactly one offline $p_T > 25 \text{ GeV } \mu$.
- Four offline $p_T > 30 \text{ GeV}$ jets are required with at least 2 b -tags.
- To enhance the $t\bar{t}$ purity, $E_T^{miss} > 20 \text{ GeV}$ and $E_T^{miss} + m_{T,W} > 60 \text{ GeV}$, where $m_{T,W} = \sqrt{2p_{T,\ell}E_T^{miss}(1 - \cos \Delta\phi_{\ell\nu})}$ is the transverse mass between the lepton and the neutrino 4-vector approximated by E_T^{miss} .

The efficiency in data (mc), $\varepsilon_{\text{data}}$ (ε_{MC}), is how often the trigger fired given the p_T . These are measured both for the L1 and HLT triggers as a function of the n offline jet p_T s, ordered by p_T . The ratio of these efficiencies is taken as the SF, and Figure 10.5 shows these SFs for the 2017 2b1j trigger. SFs are applied to the MC signals, and the error bar on the measurement is taken as an analysis systematic.

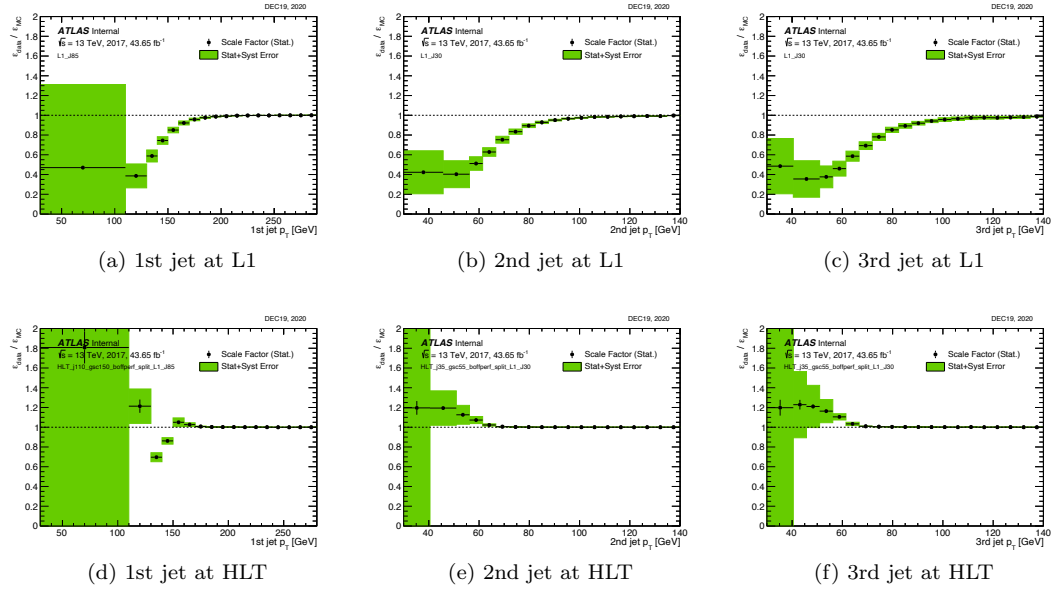


Figure 10.5: Online jet kinematic scale factors of 2b1j trigger as a function of offline jet p_T in 2017. Vertical error bars include statistical uncertainties on the data, while the green bands correspond to the quadrature sum of statistical and systematic uncertainties.

10.2 Analysis jets

10.2.1 Jets

This analysis uses particle flow jets in the fully resolved state with $R = 0.4$ jets. To suppress the contribution of jets formed by pile-up processes, jets are required to pass the JVT cut as described in Chapter 5.4. Jets produced by cosmic-rays, beam-induced background, and out-of-time pileup are reduced by imposing a set of quality criteria on variables characterizing the jet profile [150]. This analysis uses PFlow jets with $|\eta| < 4.5$ and p_T down to 30 GeV. Since the Jet/ETMiss group provides calibrations down to 20 GeV, jets with a lower p_T cutoff were studied, but not found to improve the analysis's sensitivity due to the exponential increase of multi-jet background.

Jets are separated into two groups based on their kinematics:

Central jets: $|\eta| < 2.5$, $p_T \geq 40$ GeV – jets used for triggering and the Higgs-candidates;

Forward jets: $|\eta| \geq 2.5$, $p_T \geq 30$ GeV – extra jets which improve the acceptance of jets produced in the vector boson fusion production process.

10.2.2 b -jets

For the b -tagging working-point optimization, both the MV2c10 and DL1r algorithms used dedicated trainings on the newly recommended PFlow jet collection [115]. Figure ?? shows the ggF SM signal and expected background in the $4b$ analysis for the MV2 and DL1r taggers retrained on the PFlow jet collection for the range of available working points.¹ The improvement of DL1r relative to MV2 is from the RNN tagger described in Chapter 6.4.1. The corresponding significance improvement with respect to the older tagger is summarized in Table 10.3. The higher background rejection of DL1r allowed for a loosening of the b -tagging working point from 70% to 77% with a corresponding 10% improvement in the stat-only ggF SM limits.

Tagger	$\sqrt{\sum_i s_i^2 / b_i}$	Improvement w/r.t. old tagger
MV2 70% WP (old: trained on EMTopo)	0.102	–
MV2 70% WP	0.116	14%
DL1r 70% WP	0.125	23%
DL1r 77% WP	0.128	25%

Table 10.3: Improvements from the newly retrained taggers compared to the older tagger on the EMTopo collection. Highlighted is the improvements from the optimized tagger and WP.

The other analyses saw similar improvements with moving to the new tagger and the 77% working point, and all the channels used this working point for ease in the subsequent combination. Since the majority of the HH analyses include b -jets in the final state, other channels are vetoing events with three DL1r b -tags at the 77% WP jet in the combination. This conservatively vetoed the $4b$ events, allowing us to explore a $3b$ analysis category (with the study in Appendix D).

10.2.3 b -jet corrections

The central jet calibrations focus on corrections for light-quark and gluon-initiated jets, and systematically underestimate the energy of b -jet as shown by the blue line in Figure 10.6. This is due to two main effects:

¹This b -tagging optimization study was done with an older analysis configuration with a different Higgs pairing algorithm, a different center for the signal region, and only a single training for the background estimate.

1. When the b decays semi-leptonically with a $W \rightarrow \mu\nu_\mu$ interaction in the cascade (which happens 21% of the time [106])
 - the neutrino energy is invisible in the jet reconstruction, and
 - the muonic energy is only partially accounted for in the jet’s energy estimate since the muon (μ) is not stopped in the calorimeter.
2. The b -jet fragmentation is wider than that of the corresponding light-jets², meaning fewer final state hadrons from the b -quark fragmentation are included in jet clustering (“out-of-cone” effect).

The blue line in Figure 10.6 shows this underestimate of the b -jet p_T with the standard jet calibrations. To correct for these effects, the HH analyses use a μ -in-jet + p_T -reco correction, which is shown in the yellow line.

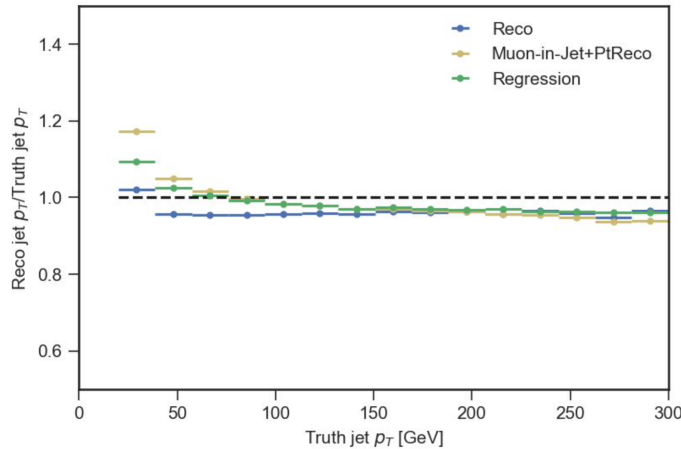


Figure 10.6: Illustration of the standard jet correction (blue), the μ -in-jet + p_T -reco used by the HH analyses (yellow), and a NN based regression (green) [151].

μ -in-jet

A search for a μ is performed in a variable radius cone $\Delta R(\mu, \text{jet}) < \min(0.4, 0.04 + 10/p_T^\mu \text{ GeV})$ ³ from the jet axis to account for the increasingly collimated decay products of more energetic jets. If a μ is identified at the medium working point with $p_T > 4 \text{ GeV}$, $|\eta| < 2.5$ is within this ΔR cone of the jet axis, its 4-vector is added to that of the jet. If there are multiple μ s passing the above criteria, only the μ closest to the jet-axis is added. Then the expected energy that the μ lost in the calorimeter is subtracted as this contribution was already included in the jet energy estimate.

²The tracks from the heavy flavor decay carry $\approx 70\%$ of the b -quark’s momentum. This makes the fragmentation tracks correspondingly softer. The opening angle of the decay products of a particle of mass m follows the relation $\Delta R \sim 2m/p_T$, giving softer fragmentation tracks in b -jets have a wider opening angle.

³The min avoids adding a μ farther away from the jet axis than the jet clustering distance parameter.

p_T -reco

This step accounts for the missing neutrino energy and out-of-cone effects that the standard jet calibrations don't capture. This correction factor is derived in $t\bar{t}$ events to correct the reconstructed p_T of the b -jets in logarithmic bins of the truth jet p_T . Since the correction is larger for b -jets decaying semi-leptonically, these correction factors are derived separately for b -jets with and without a μ .

Figure 10.7 illustrates the improvement achieved by the b -jet corrections in the m_{H1} and m_{H2} resolution. The blue line shows the reconstructed Higgs Candidate (HC) masses from the two b -jets with the standard jet calibration, while the red line shows the reconstructed Higgs masses with the μ -in-jet + p_T -reco b -jet correction applied. This increases the central value of the reconstructed HCs, with the subleading (in p_T) HC central value shifting from 114.5 GeV to 120.2 GeV. This also improves the resolution on the reconstructed Higgses by sharpening the signal peaks. The reconstructed width improves by 5.2% for the leading HC and 18.5% for the subleading HC. With this improvement in the reconstructed central value of the Higgs mass and resolution, we optimized the signal region to improve the signal efficiency while decreasing the background.

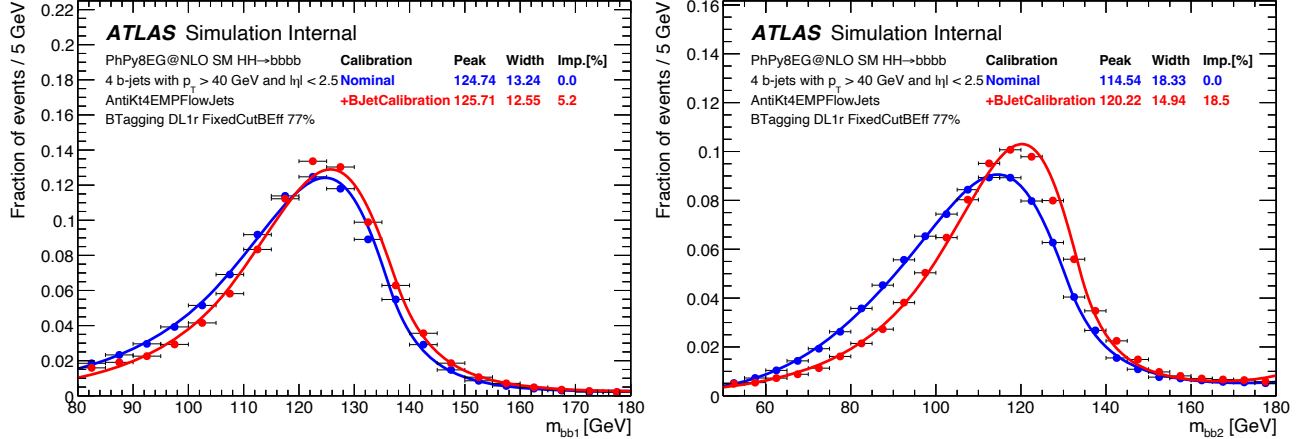


Figure 10.7: Comparisons of m_{H1} and m_{H2} distributions before the b -jet corrections (blue) and after the b -jet corrections (red). These distributions are fitted using Bukin function, and the peak, the peak resolution, and the relative improvement are shown in the legend.

10.3 Event reconstruction

10.3.1 ggF and VBF Channel orthogonalization

This analysis targets the ggF and VBF production modes two channels. In the combined set of the central and forward analysis jets that are not b -tagged, the pair of jets with the highest invariant mass (or largest m_{jj}) are the candidate VBF jets. These VBF jets are used to define three cuts to decide whether this event has VBF-like kinematics. The VBF jets need to have an $m_{jj} > 1000$ GeV and a pseudo-rapidity gap $\Delta\eta_{jj} > 3$. Additionally, the four-vector sum of the VBF jets and the four jets of the HH system needs to have $p_T < 65$ GeV. Figure 10.8 shows distribution for these three variables and the VBF channel cuts. The dashed pink line show the SM ggF signal, and the solid lines show the two VBF signals (pink for VBF SM and navy for VBF $\kappa_{2V}=0$). If an event passes these three cuts, it goes into the “VBF channel”, and events failing any of these cuts are kept in the “ggF channel”.

Only $\approx 1.5\%$ of the ggF signals go into the VBF category. Since the VBF cuts were optimized for the significance (not the signal yield), only 13% – 17% of the VBF signal events go in the VBF channel (see Table 10.5 for the SM, $\kappa_\lambda=10$, and $\kappa_{2V}=0$ VBF signals).

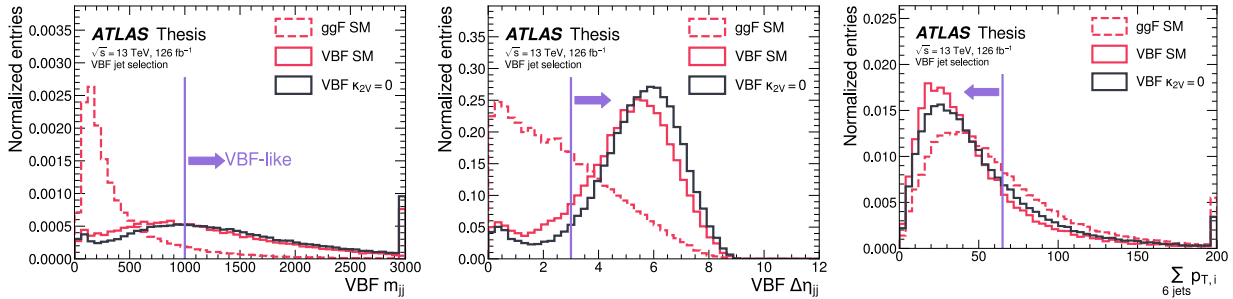


Figure 10.8: Illustration of the variables to implement the ggF / VBF channel orthogonalization. The purple lines and arrows indicate the three cuts for an event to pass the VBF selection.

10.3.2 Higgs candidate reconstruction

Higgs Candidate Jets

The HH system is reconstructed with four central b -tagged jets. If there are more than four b -tags, the four leading (in p_T) jets are chosen. The Higgs *jet selection accuracy* is the probability that all four of the chosen b -jets are within $\Delta R < 0.3$ of the truth b -quarks coming from the Higgs decays. The jet selection accuracy is shown in Figure 10.9 for the κ_λ and κ_{2V} variations. For the ggF SM signal with the ggF selection this accuracy is 74% with %-level variations across the κ_λ values of interest. This accuracy loss is dominated by events where one of the b -quarks is out of acceptance.

The 4b VBF selection has an average b -quark selection accuracy of 85% and 90% for the respective κ_λ and κ_{2V} signal samples. Since the b -tagging efficiency increases with the b -jet p_T (as shown in Figure 6.22(a)), the harder VBF signals have higher b -tagging efficiency which gives a higher accuracy for tagging the b -jets from the Higgs decays. Figure 10.10 shows the truth m_{HH} distributions for the case where the correct jets were selected (solid lines) and where one of the selected jets is not matched to a b -quark from the Higgs decay (dashed). This is shown both for the ggF selection (left) and the VBF selection (right) at a few signal points. The correct jets are less likely to be selected for lower m_{HH} . Although the signal shapes for the κ_λ and κ_{2V} variations are quite different, the ratios in the sub-panels show the accuracy differences are entirely due to the different m_{HH} distributions.

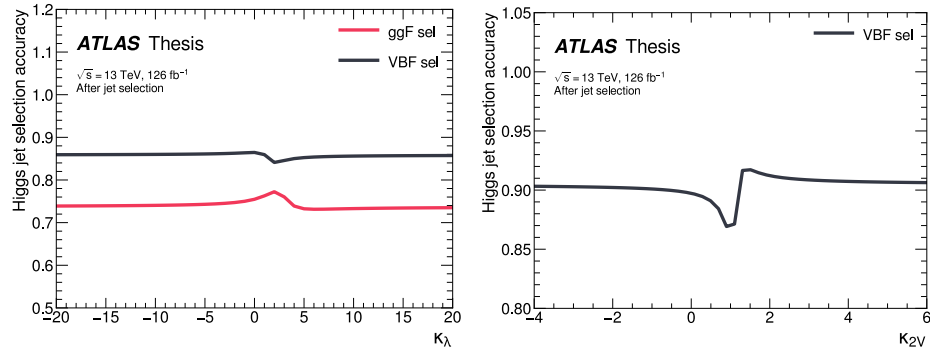


Figure 10.9: The jet selection accuracy as a function of κ_λ (left) and κ_{2V} (right).

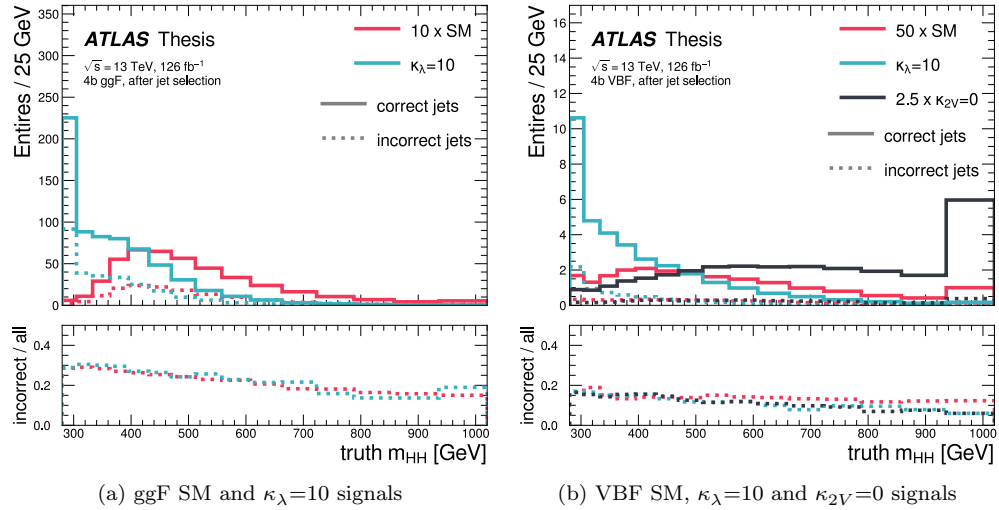


Figure 10.10: Truth m_{HH} distributions for correctly and incorrectly selected jets, for the ggF (left) and VBF (right) signals.

Higgs candidate pairing

These four jets are then paired to reconstruct Higgs Candidates (HCs). With four selected jets, three pairings are possible, as sketched in Figure 10.11. The angular separation of decay products a heavy resonance are expected to follow:

$$\Delta R_{jj} \lesssim \frac{2m}{p_T},$$

where m and p_T are the mass and p_T of the resonance. Since higher p_T objects have more collimated daughters, to define the pairing, we chose the pairing that had the smallest ΔR_{jj} between the constituent jets for the Higgs Candidate with the higher p_T . The **correct pairing** is defined with generator-level information by requiring the b -quarks which have the same parent barcode ID in the truth record.

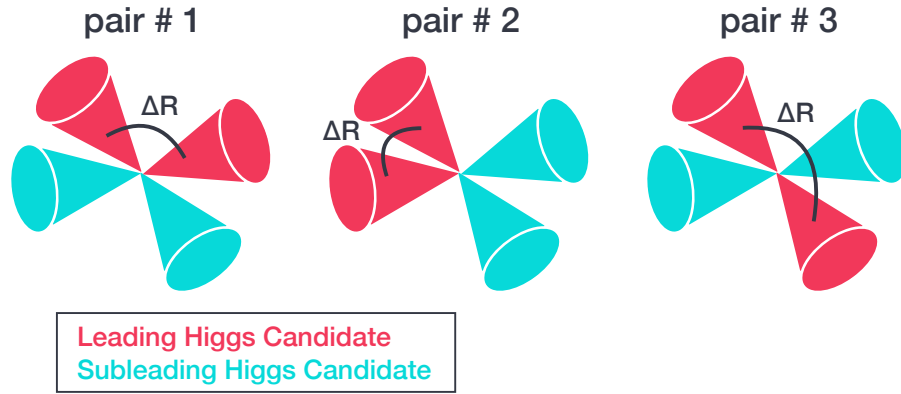


Figure 10.11: The three possible pairing permutations of the four HH jets into the two Higgs candidates. The opening angles between the jets in the leading Higgs Candidate are indicated, so pair number 2 is the selected pairing.

The pairing accuracy is the fraction of correctly paired events among the events where the four Higgs-decayed jets are correctly selected by the jet selection, which decouples the pairing accuracy from the jet selection accuracy. The pairing accuracy is shown in Figure 10.12 as a function of κ_λ and κ_{2V} , and in Figure 10.13 as a function of m_{HH} . Signals with harder p_T Higgses have more collimated jet pairs, leading to a corresponding accuracy loss for m_{HH} events below 450 GeV (Figure 10.13). This means that the κ_λ variations have a lower accuracy than the SM, and the κ_{2V} variations have higher pairing accuracies. We considered other pairing strategies that improved the low m_{HH} accuracy. However, algorithms with improved low m_{HH} efficiency also sculpted the background to peak inside the SR, making it more challenging to derive a trustworthy background estimate. We chose this min ΔR pairing algorithm to reject the low m_{HH} backgrounds and optimize our sensitivity for the harder signals. In Chapter ?? further motivates this choice in the discussion comparing the strengths of the main HH channels.

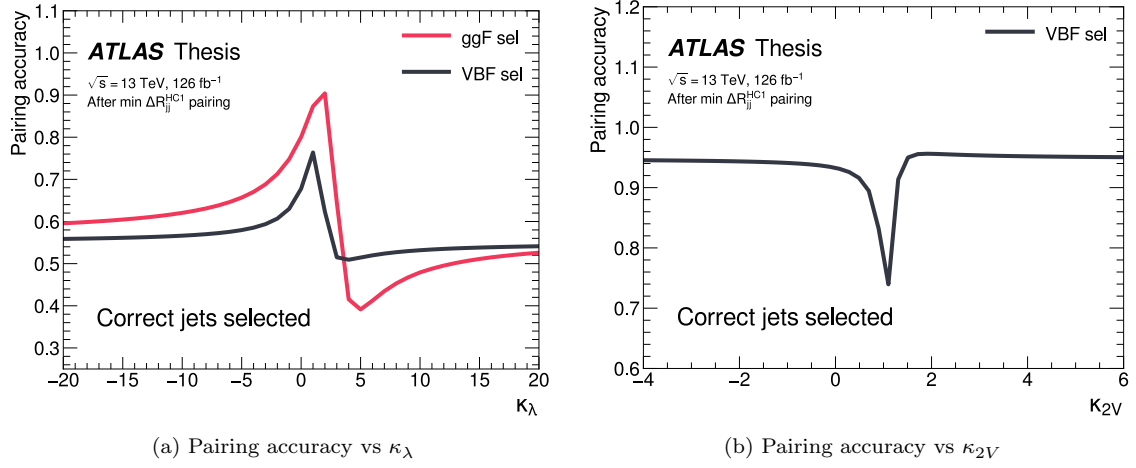


Figure 10.12: The pairing accuracy as a function of κ_λ and κ_{2V} , given that the correct jets have been selected.

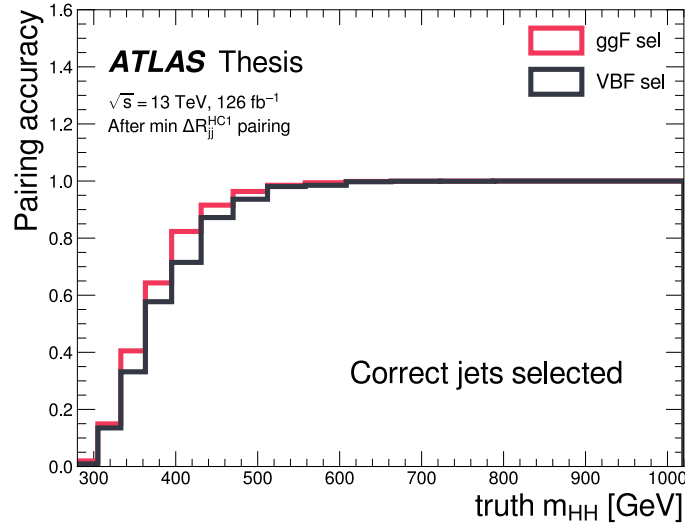


Figure 10.13: Pairing accuracy as a function of truth m_{HH} , given that the correct jets have been selected. The ggF selection accuracy is derived from the ggF SM sample, and the VBF selection accuracy is derived from the VBF $\kappa_{2V}=0$ sample.

10.3.3 Background Reduction

$\Delta\eta_{HH}$

The pseudo-rapidity difference between the reconstructed HCs, $\Delta\eta_{HH}$, also helps with signal versus background discrimination. The triangle diagram for the ggF signals has an s-channel propagator so the production rate for the ggF SM and $\kappa_\lambda=10$ signals decreases as $\Delta\eta_{HH}$ increases (Figure 10.14(a)). To compare this to the QCD background we show the blinded 4b data (events in the 4b SR, Eq. ??, are not shown). The QCD background processes have more t -channel production, resulting in a less steeply falling $\Delta\eta_{HH}$ distribution [152]. The ggF channel applies a cut of $\Delta\eta_{HH} < 1.5$ to reject this multi-jet background. Figure 10.14(b) shows that the SM VBF signal has a flat $\Delta\eta_{HH}$ distribution out till $\Delta\eta_{HH} = 3$, and the majority of the SM signal is at $\Delta\eta_{HH} > 1.5$. The BSM κ_λ and κ_{2V} variations peak at lower $\Delta\eta_{HH}$, so to retain sensitivity to both of these signals, the VBF channel does *not* apply the $\Delta\eta_{HH}$ cut. $\Delta\eta_{HH}$ is a useful categorization variable for both the ggF and VBF channels, as will be discussed in Chapter 10.5.

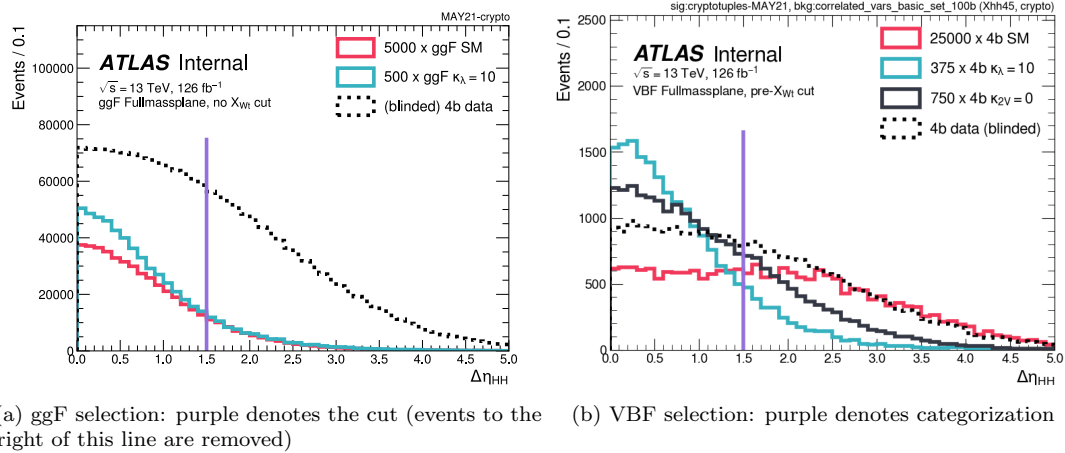


Figure 10.14: The $\Delta\eta_{HH}$ distribution for HH MC and blinded 4b data. The solid purple line indicates the $\Delta\eta_{HH} < 1.5$ cut applied in the ggF selection (right). This cut is not applied for the VBF channel, but denotes the categorization.

Top veto

A top veto suppresses backgrounds from hadronic top-quark decays. W candidates are formed from any pair of central jets in the event, and top candidates are built by pairing W candidates with b -jets that were selected for Higgs candidates. The discriminant X_{Wt} is constructed for each combination:

$$X_{Wt} = \sqrt{\left(\frac{m_W - 80.4 \text{ GeV}}{0.1 m_W}\right)^2 + \left(\frac{m_t - 172.5 \text{ GeV}}{0.1 m_t}\right)^2} \quad (10.2)$$

We take the minimum X_{Wt} is the configuration most likely to reconstruct the top, and we cut on $X_{Wt} > 1.5$. Figure 10.15 shows these X_{Wt} distributions for $t\bar{t}$ events and the benchmark signals, with both for the ggF and VBF selections, and the $t\bar{t}$ signals peak at lower X_{Wt} values.

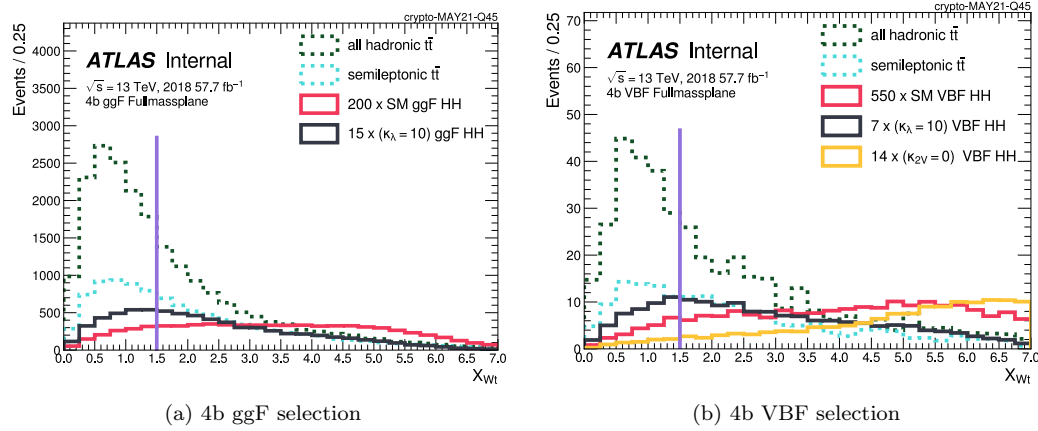


Figure 10.15: X_{Wt} distributions with the 2018 dataset for the ggF (left) and VBF (right) channels. The solid purple line indicates the $X_{Wt} > 1.5$ cut.

Signal region definition

The signal region definition uses X_{HH} :

$$X_{HH} = \sqrt{\left(\frac{m_{H1} - 124 \text{ GeV}}{0.1 m_{H1}}\right)^2 + \left(\frac{m_{H2} - 117 \text{ GeV}}{0.1 m_{H2}}\right)^2}. \quad (10.3)$$

The functional form of this variable is similar to the equation for an ellipse, except that the radius is a function of the Higgs Candidate (HC) masses to allow harsher cuts for higher HC masses where the jets' resolution is better. The values (124 GeV, 117 GeV) in Eq. 10.3 correspond to the mean reconstructed masses for correctly paired signal events. The SR is defined by $X_{HH} < 1.6$. Figure 10.16 shows the benchmark ggF and VBF signals peak nicely inside of this SR. Figure 10.17 shows X_{HH} for correctly and incorrectly paired signal events and demonstrates that $X_{HH} < 1.6$ has a high purity for correctly paired signal.

Figure 10.18 shows the (blinded) 4b and 2b data (m_{H1}, m_{H2}) “massplanes” with the ggF selection. The backgrounds for the 4b distributions will be built from reweighted 2b data using a data-driven background estimate built from kinematically similar control regions (CRs). Two control regions are defined:

- Control Region 1 (CR1): where we derive the data-driven background estimate
- Control Region 2 (CR2): where we assess an uncertainty for the extrapolation into the SR.

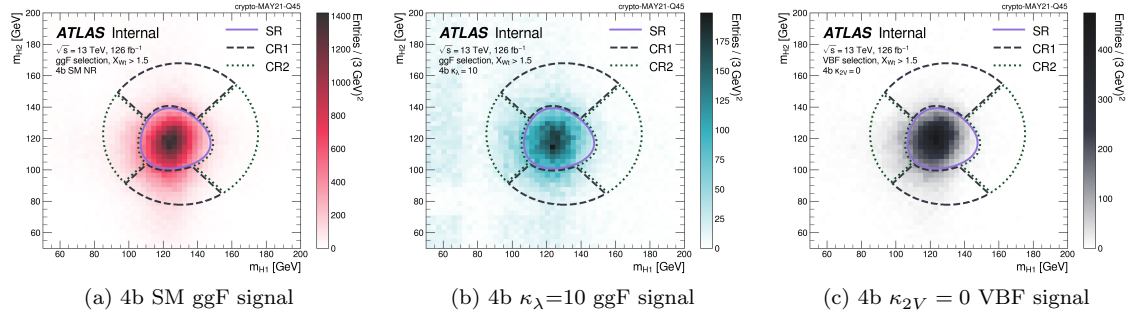
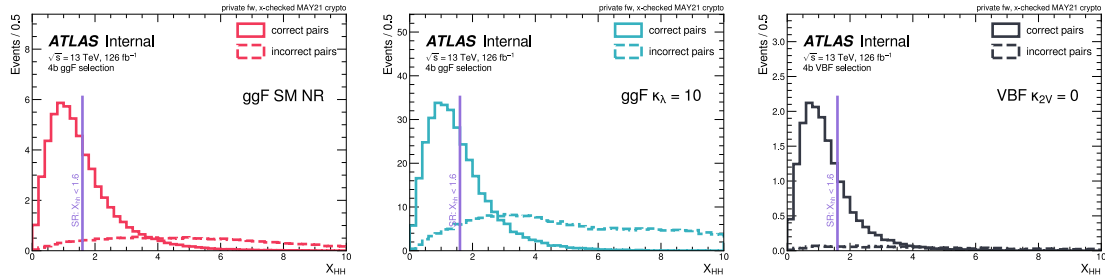


Figure 10.16: Selected Higgs Candidate signal mass planes.

Figure 10.17: Visualization of the X_{HH} distribution for correctly and incorrectly paired events with the SM ggF (left), ggF $\kappa_\lambda=10$ (center), and $\kappa_{2V}=0$ VBF (right) analysis selections. The purple line indicates the SR defining cut.

These CR1 and CR2 regions lie within

$$\text{CR boundary} \quad : \quad \sqrt{(m_{H1} - 1.05 \cdot 124 \text{ GeV})^2 + (m_{H2} - 1.05 \cdot 117 \text{ GeV})^2} < 45 \text{ GeV}. \quad (10.4)$$

The 45 GeV CR radius comes from a trade-off of having sufficient statistics to derive the background estimate while being small enough for the CR to be kinematically similar to the SR. The CR center is shifted by 5% relative to the SR center to keep the m_{H1}, m_{H2} means similar to the SR means. The CR1 and CR2 regions are defined with lines at 45° intersecting the SR. The union of the top and bottom crescents creates CR1, and the left and right crescents compose CR2. The 45° for the CR1 / CR2 definitions was chosen to minimize the non-closure of the background estimates in a validation region with fewer b -tags (the “3b1f” region, described in Chapter 11.3.3).

In Figure 10.18, there's a deficit of events for m_{H1} or m_{H2} around 80 GeV showing up as light-colored bands that stretch horizontally and vertically across the plot. The X_{Wt} cut is imposing a W -mass veto for the constructed Higgs Candidates (HCs), and Figure 10.19 shows that these bands are removed when the X_{Wt} cut is omitted.

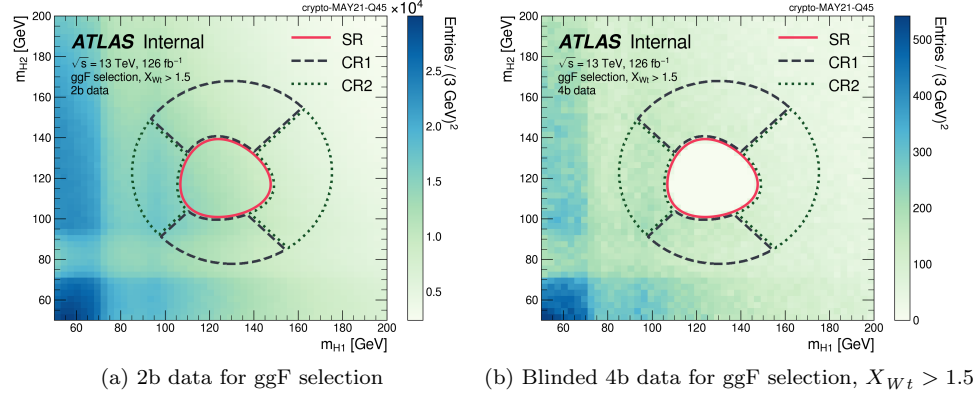


Figure 10.18: The Higgs Candidate massplanes with the ggF analysis selection.

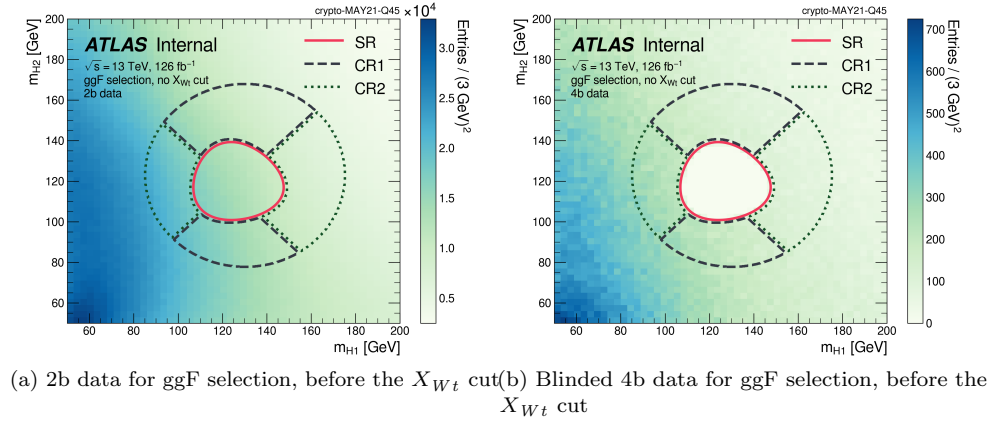


Figure 10.19: The Higgs Candidate massplanes with the ggF analysis selection, **before applying the X_{Wt} cut.**

10.4 Signal Cutoffs

The acceptance times efficiency plots for each of the analysis cuts are shown in Figure 10.20. Here we show the ggF signals in the ggF channel and the VBF signals in the VBF channel. Figure 10.20(a)

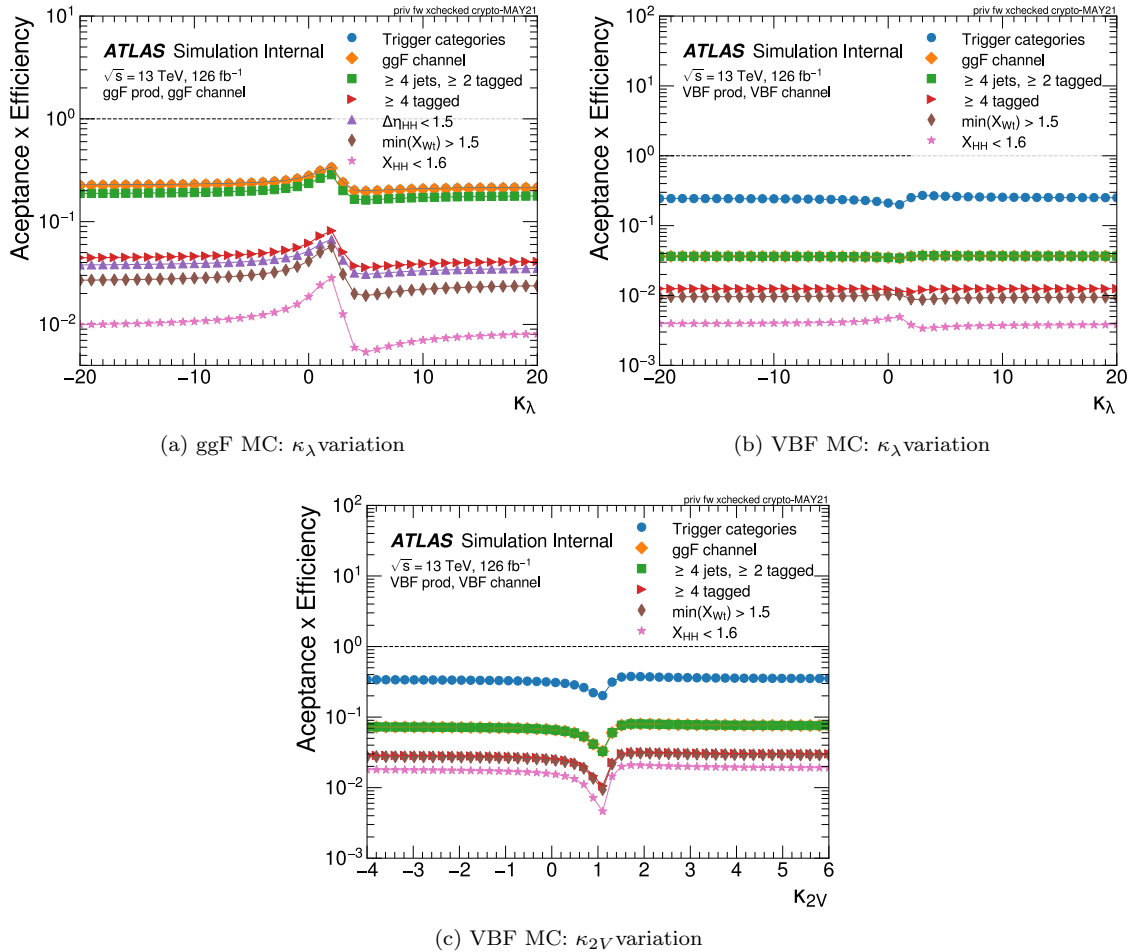


Figure 10.20: $HH \rightarrow 4b$ acceptance times efficiency plots.

shows that the acceptance x efficiency for the ggF signals is much lower for the κ_λ variations, and is lowest for the softest $\kappa_\lambda = 5$ value. For the $4b$ non-resonant analyses, we optimized for the SM sensitivity, as high m_{HH} is the strength of the $4b$ analysis. Figure 10.21 illustrates this by showing limits from a recent search for resonances X decaying to HH pairs for three main HH channels: $4b$, $b\bar{b}\tau^+\tau^-$, and $b\bar{b}\gamma\gamma$ [153]. Although the $4b$ channel dominates the combination's sensitivity for the high resonance masses ($m_{HH} > 700$ GeV) for the lowest resonance masses ($m_{HH} < 350$ GeV) the $4b$ channel is much less sensitive than the other two channels. This convinced us that even if we

optimized for low m_{HH} we still wouldn't have an impact in the combination for low m_{HH} relative to $bb\gamma\gamma$.

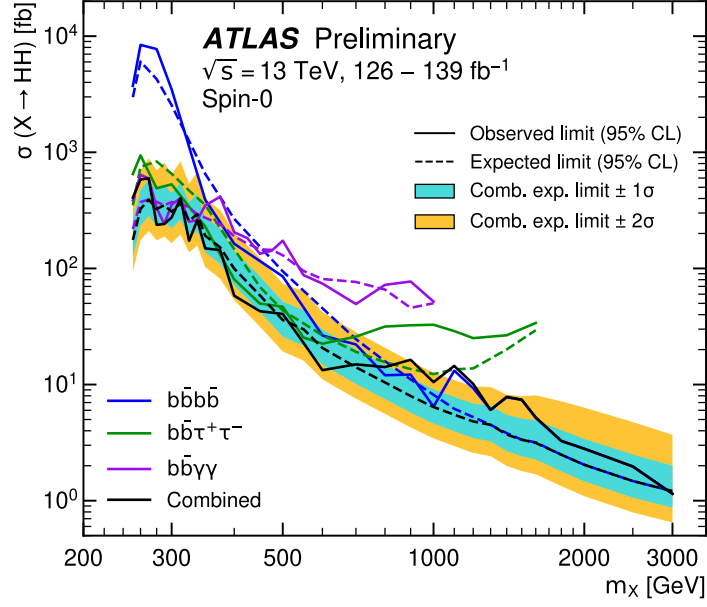


Figure 10.21: Impact of the HH channels in the combination for the resonant scalar mass m_X search [153].

To illustrate how this translates into the characteristics of the κ_λ variations, Figure 10.22 shows the impact of the analysis cuts on for a few representative signals: SM, $\kappa_\lambda=2$, and $\kappa_\lambda=10$. For the $\kappa_\lambda=2$ signal, we don't reconstruct any of the events at low m_{HH} of the double peak from the destructive interference between the two diagrams. For the $\kappa_\lambda=10$ signal, we also don't reconstruct most of the low m_{HH} events from the enhanced triangle diagram.

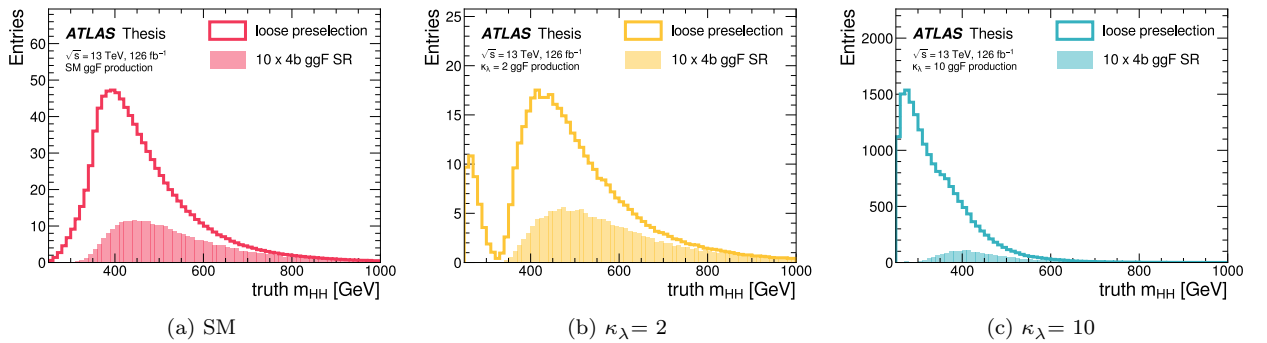


Figure 10.22: Impact of the NR analysis selection for selected ggF signals.

Table 10.4 gives the cutflows for the data, and cutflows for the ggF and VBF benchmark signals are in Appendix C.2.

	Yield	Yield / Pre-selection	Yield / Prior cut
Initial (Unweighted for MC)	1.59e+10	-	-
Pass NTuple Preselection	5.697e+08	1	-
Trigger	2.807e+08	0.4927	0.4927
Trigger Buckets	2.49e+08	0.4371	0.8873
ggF channel	2.457e+08	0.4314	0.9868
≥ 4 central jets, ≥ 2 b -tags	1.806e+08	0.317	0.7349
≥ 4 b -tags	1.886e+06	0.003311	0.01045
$ \Delta\eta_{hh} < 1.5$	1.032e+06	0.001811	0.5469
Top Veto	7.506e+05	0.001318	0.7276
Signal Region	1.617e+04	2.839e-05	0.02154
Control Region 2	3.067e+04	5.383e-05	0.04085
Control Region 1	3.204e+04	5.625e-05	0.04268

(a) 4b data (ggF channel)

	Yield	Yield / Pre-selection	Yield / Prior cut
Initial (Unweighted for MC)	1.59e+10	-	-
Pass NTuple Preselection	5.697e+08	1	-
Trigger	2.807e+08	0.4927	0.4927
Trigger Buckets	2.49e+08	0.4371	0.8873
VBF channel	3.295e+06	0.005784	0.01323
≥ 4 central jets, ≥ 2 b -tags	3.157e+06	0.005543	0.9583
≥ 4 b -tags	2.711e+04	4.759e-05	0.008586
$ \Delta\eta_{hh} < 1.5$	2.711e+04	4.759e-05	1
Top Veto	2.175e+04	3.818e-05	0.8024
Signal Region	502	8.812e-07	0.02308
Control Region 2	906	1.59e-06	0.04165
Control Region 1	947	1.662e-06	0.04354

(b) 4b data (VBF channel)

Table 10.4: 2016-18 data yields at each step in the analysis event selection for 4b events in the ggF (top) and VBF (bottom) channels, alongside the ratio of each yield to the initial yield and to the yield for the previous cut. Highlighted are the yields in the 4b SRs.

Table 10.5 shows the 4b SR yields for the ggF and VBF signals in the ggF and VBF channel SRs. Approximately 1–2% of the ggF signal in the VBF channel SR, demonstrating that VBF veto does not dilute the ggF sensitivity. Only 30% - 60% of the VBF signal yield lies in the VBF channel’s SR, but the VBF channel has a much lower background, therefore better significance for the VBF signals.

	ggF channel yield	VBF channel yield	% in VBF channel
ggF signals			
SM	29.1	0.4833	1.63
$\kappa_\lambda = 10$	182.7	3.099	1.67
VBF signals			
SM	0.2352	0.3265	58.1
$\kappa_\lambda = 10$	25.19	14.24	36.2
$\kappa_{2V} = 0$	22.97	17.29	42.9
4b data	16,170	502	3.01

Table 10.5: Yields in the 4b ggF and VBF SRs for VBF and ggF HH signals normalized to 126.1fb^{-1} .

10.5 Analysis Categories

Previous iterations of this analysis used the invariant mass of the HH system (m_{HH}) as the discriminating variable for extracting limits on the signal strength [116, 117, 154]. A multi-variate algorithm (MVA) such as a BDT or NN would provide additional discrimination power. But with our fully data-driven background estimate, we were concerned if we could model and validate all of the variable correlations exploited by the MVA. As a compromise, a number of discriminating variables are used to define extra categories of S/\sqrt{B} purity, as this gave a finite number of correlations to check for the background estimate, which we were able to in several validation regions (described in Chapter 11.3). Since the m_{HH} distribution is steeply falling, a logarithmic binning is used to keep the high m_{HH} bins while keeping reasonable statistical uncertainties.

ggF categories

The ggF channels are categorized in $\Delta\eta_{HH}$ and X_{HH} , with the signal and background distributions in the SR shown in Figure 10.23. The upper cutoffs on these histograms correspond to previously defined ggF channel cuts of $\Delta\eta_{HH} < 1.5$ for the QCD background rejection and $X_{HH} < 1.6$ for the SR definition. Since the background $\Delta\eta_{HH}$ distribution is flat, three equally spaced $\Delta\eta_{HH}$ bins were chosen between 0 and 1.5. Additionally, two X_{HH} bins were defined, with the boundary of 0.95 optimizing the SM S/\sqrt{B} significance.

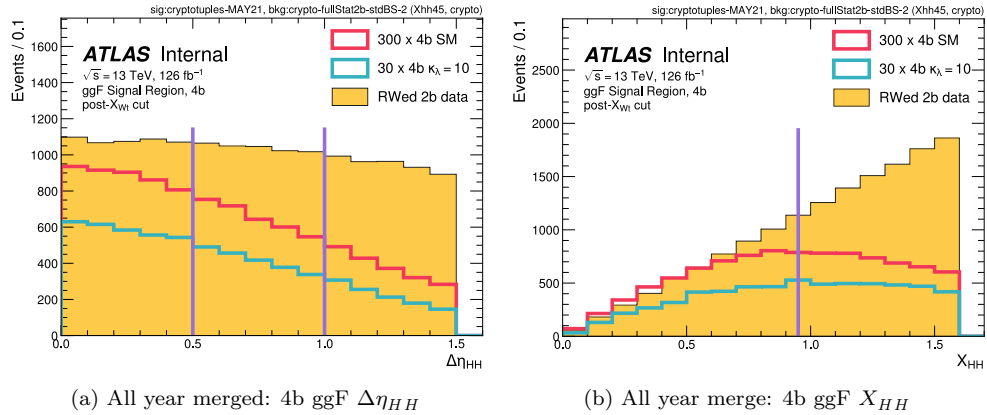


Figure 10.23: Distributions of the categorization variables for the ggF channel. To visualize the signals they are scaled by $\alpha = 300$ and 30 for the ggF SM and $\kappa_\lambda = 10$ signals, respectively.

The ggF background estimate is derived separately for each years, and the corresponding 4b ggF discriminants are shown in Figure 10.24, with the SM and $\kappa_\lambda = 10$ signals overlaid. The sub-panels on these plots show the S/\sqrt{B} significance. The signal peaks for the lower $\Delta\eta_{HH}$, X_{HH} values, and these are the high purity categories that drive the significance.

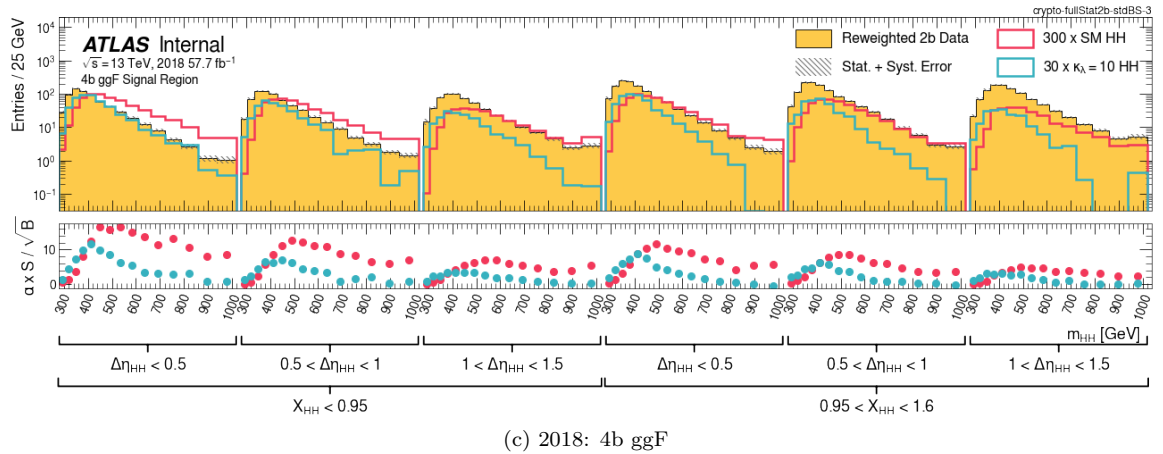
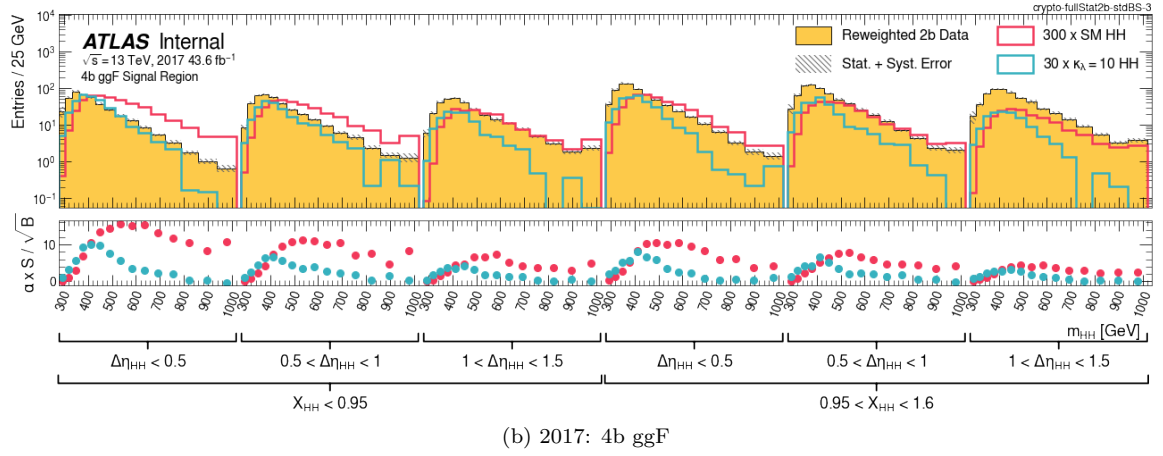
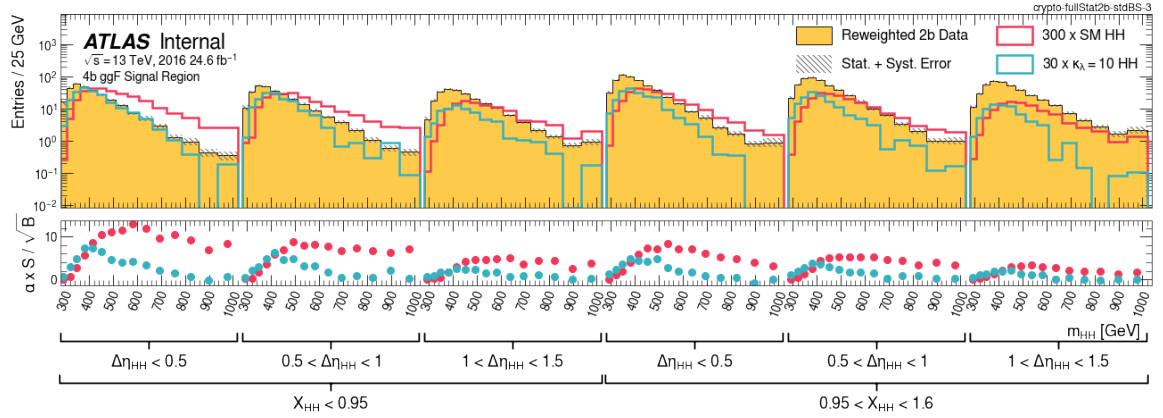


Figure 10.24: 4b ggF background and selected signal histograms for 2016, 2017, and 2018 with the proposed binning and categorization. To visualize the signals they are scaled by $\alpha = 300$ and 30 for the SM NR and $\kappa_\lambda=10$ signals, respectively.

VBF categories

The VBF channel uses two $\Delta\eta_{HH}$ categories with a boundary at $\Delta\eta_{HH} = 1.5$ that maximized the significance. Figure 10.25 shows the $\Delta\eta_{HH}$ background distribution along with three VBF signals: $\kappa_\lambda = 10$, $\kappa_{2V} = 0$, and the SM. The $\Delta\eta_{HH}$ distribution of the non-SM couplings peaks close to $\Delta\eta_{HH} = 0$, while the distribution corresponding to the SM prediction peaks at approximately $\Delta\eta_{HH} = 2$. The low $\Delta\eta_{HH} < 1.5$ category drives the sensitivity to the non-SM couplings, while the $\Delta\eta_{HH} \geq 1.5$ category is targetted for the SM sensitivity.

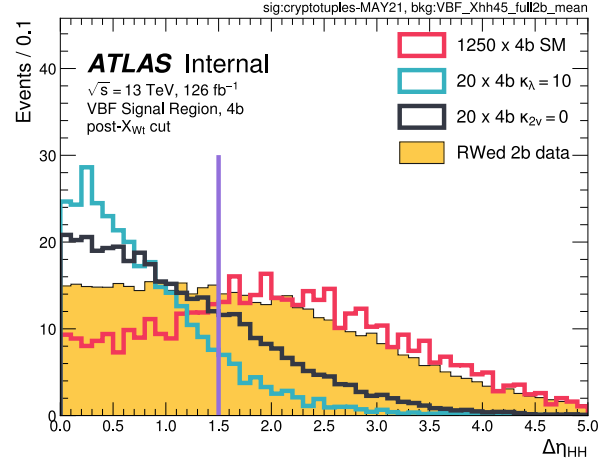


Figure 10.25: $\Delta\eta_{HH}$ categorization variable for the VBF channel. The purple line indicates the $\Delta\eta_{HH} = 1.5$ category boundary. To visualize these signals they are scaled by $\alpha = 1250$ and 20, and 20 for the VBF SM, $\kappa_\lambda = 10$, and $\kappa_{2V} = 0$ signals, respectively.

Figure 10.26 shows the m_{HH} in these categories with the same three benchmark signals and the background prediction. The significance in sub-panel demonstrates how the high $\Delta\eta_{HH}$ bin drives the SM sensitivity, while the low $\Delta\eta_{HH}$ bin drives the BSM couplings sensitivity. Due to the low significance below 400 GeV and difficulties in the low m_{HH} modelling, the bins of $m_{HH} < 400$ GeV are not included in the final fit for either category.

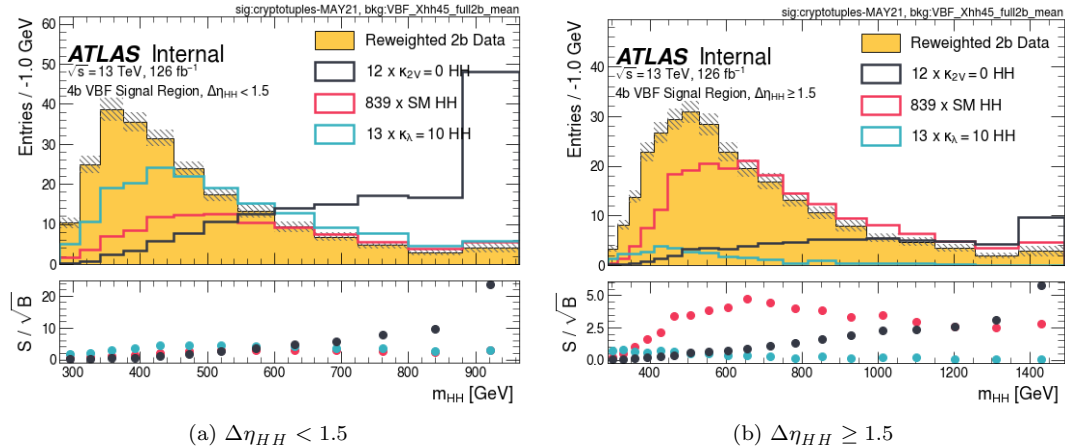


Figure 10.26: 4b VBF background and selected signal histograms for 2016, 2017, and 2018 with the proposed binning and categorization. To visualize these signals they are scaled by $\alpha = 1250$ and 20, and 20 for the VBF SM, $\kappa_\lambda = 10$, and $\kappa_{2V} = 0$ signals, respectively.

11

Background estimation

This IS the analysis.

– Georges Aad (editorial board chair for collaboration internal review)

Although the 4b final state is tantalizingly attractive due to its maximal HH branching ratio, the fully hadronic final state also means coping with larger background rates due to the plethora of multi-jet events produced at a hadron collider, processes that are notoriously difficult to simulate from first principles.

Figure 11.1 emphasizes the degree of discrepancy for the MC at estimating the data yield in the background enriched control region, CR1. The simulated $t\bar{t}$ and Pythia 8.2 QCD processes fail to accurately predict the data, with errors up to 50% for high m_{HH} . This motivates us to resort to other methods to form a background prediction in the signal region, and accurately estimating this background and assigning an appropriate error bar is the main challenge of this analysis.

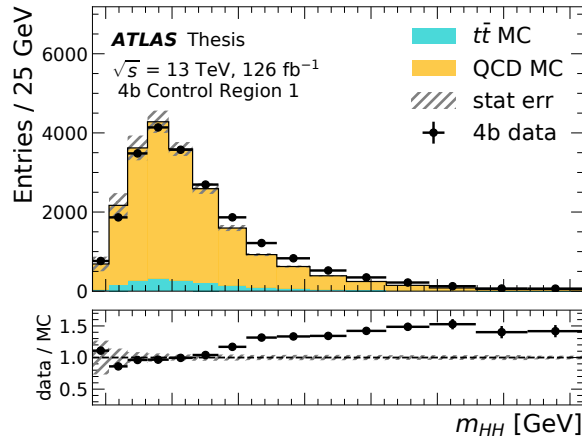


Figure 11.1: The number of events per bin as a function of di-Higgs candidate mass for the data and the Monte Carlo estimates for multi-jet and top pair-production process with 4 b -jets in Control Region 1 (defined in Chapter 10.3.3).

Chapter 11.1 overviews the nominal background estimate strategy, while Chapter 11.2 describes the expected composition of the background. Chapter 11.3 enumerates the errors assessed on this estimate, and Chapter 11.4 concludes with tests of this prescription in different validation regions.

11.1 Reweighting overview

The background estimate reweights distributions from a lower b -tag region into a higher b -tag region by deriving reweighting maps in dedicated control regions. The lower b -tagged region consists of events with exactly two b -tags (“ $2b$ ” events), while the higher tag region contains events with four or more b -tags (“ $4b$ ” events). For the $2b$ events, the two leading in p_T non- b tagged jets are taken for the other jets, and the rest of the analysis cuts are applied.

Figure 11.2 shows the regions relevant in this Chapter (with the equations defined in Chapter ??) The *nominal estimate* derives the reweighting maps are derived in a kinematically similar control region, the navy “Control Region 1” (CR 1) upper and lower crescents in Figure 11.2. Extrapolating the nominal estimate to apply it in the $2b$ SR estimates the $4b$ SR data, like a generalized ABCD method. An error on this estimate from the choice of training region is evaluated by rederiving the reweighting maps in the alternative “Control Region 2” (CR 2), shown in the turquoise left and right crescents. The difference between the CR1 and CR2 predictions gives an error bar, which will be described in Chapter 11.3.2.

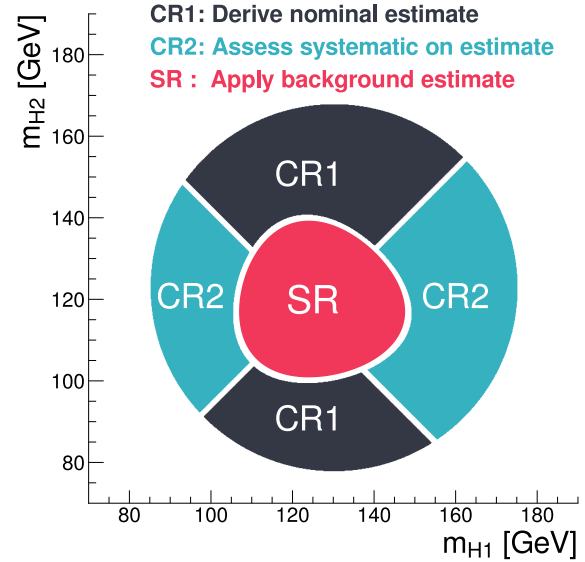


Figure 11.2: The regions in the $4b$ analysis.

The “reweighting” consists of deriving the maps $w(x)$ such that

$$p_{4b}(x) = w(x) \cdot p_{2b}(x), \quad (11.1)$$

where p_{4b} , p_{2b} denotes the probability distribution over the $4b$ and $2b$ events, respectively, and x is a set of features characterizing the kinematics of the event.

In past iterations of the analysis, this reweighting function $w(x)$ was determined by iteratively correcting each of the features in x by fitting the $4b/2b$ ratios of the 1d histograms [116]. This sequential reweighting of the x_i features fails to fully model for the correlations between the x_i features, which led to a large mismodeling of the low m_{HH} events. This approach also didn’t scale well for increasing the number of features, as x was just a 6-dimensional vector.

An alternative approach is to use the “Density Ratio Trick” where a classifier $D(x)$ can be used to discriminate between the $2b$ and $4b$ data. The $4b$ events are labeled as class 1 and the $2b$ events as class 0, and a non-linear sigmoid function $\sigma(z) = \frac{1}{1+\exp(-z)}$ ensures the classifier outputs are also

between 0 and 1. The parameters of $D(x)$ can be found by minimizing the logistic loss [155]:

$$\mathcal{L} = -\mathbb{E}_{x \sim p_{4b}} [\log D(x)] - \mathbb{E}_{x \sim p_{2b}} [1 - \log D(x)]. \quad (11.2)$$

The minimum of this loss function is derived in Appendix E.1 and is:

$$D(x) = \frac{1}{1 + \frac{p_{4b}(x)}{p_{2b}(x)}}.$$

This allows us to solve for the reweighting maps given the classifier output.

$$w(x) = \frac{p_{4b}(x)}{p_{2b}(x)} = \frac{D(x)}{1 - D(x)}$$

By using classifiers to implement the reweighting, we can avoid the curse of dimensionality by using more modern ML classifiers which scale favorably with higher dimensional inputs, such as BDTs or NNs. This innovation was used in the $HH(\rightarrow 4b) + E_T^{\text{miss}}$ SUSY analysis where a BDT classifier was used to implement the background reweighting [156].

For this analysis, instead of using a NN directly as a classifier, we instead learn $Q(x) = \log w(x)$, as proposed in [157]. We do this by minimizing the loss function:

$$\mathcal{L}[Q] = \mathbb{E}_{x \sim p_{2b}} \left[\exp \left(\frac{1}{2} Q(x) \right) \right] + \mathbb{E}_{x \sim p_{4b}} \left[\exp \left(-\frac{1}{2} Q(x) \right) \right] \quad (11.3)$$

which results in a model learning $Q^*(x) = \log w^*(x)$ (proof in Appendix E.2).

This solution also avoids the curse of dimensionality while accounting for correlations, but there is now no non-linearity on the output classifier score as $Q(x) \in \mathbb{R}$. This avoids the saturating gradients problem inherent with using sigmoids in deep learning solutions. Also, learning the log of the likelihood ratio naturally enforces positive weights as $w(x) = e^{Q(x)}$. Experiments in [157] show that this exponential loss gives (slightly) better performance than the logistic loss – and our team saw similar results in our analysis optimization.

The ggF (VBF) networks have 3 hidden layers with 50 (20) hidden units each and use ReLU [158] non-linearities. The training is done using stochastic gradient descent with the adam [112] optimizer with a learning rate of 0.001 and a batch size of 1024 events. 20% of the CR1 training events are held out as a validation set, and the optimization continues until the loss on the validation set hasn't improved in the last 15 epochs¹.

The input variables for the ggF and VBF reweightings are delineated in Table 11.1. While the sequential reweighting from [116] was limited to $x \in \mathbb{R}^6$, our ML solution could handle $x \in \mathbb{R}^{12}$ for ggF and $x \in \mathbb{R}^9$ for VBF. These inputs were optimized to best characterize the differences between the $2b$ and $4b$ data. One sources of the heavy flavor multi-jet background is $g \rightarrow b\bar{b}$ splitting [159],

¹An epoch is a single pass over the training dataset.

which is more likely in $4b$ events than in $2b$ events since there are more b -jets. Enumerating the b -quarks from the gluon by the indices 1, 2, the invariant mass of the gluon is $m_{12}^2 \approx 2E_1E_2(1 - \cos\theta)$. As a $b\bar{b}$ pair that are nearly collinear, the gluon will go on-shell. The smallest opening angle between jets in the event, “ ΔR_{jj} Close” is a reweighting input that identifies the gluon splitting kinematics [160]. The opening angle between the other two jets in the event is another reweighting variable, “ ΔR_{jj} Not Close”. Since different triggers are used for each year of the data, the ggF background estimate has a different background estimate derived for each of the three years. Since VBF has two orders of magnitude less statistics than ggF, it was not as sensitive to these differences, so the VBF training was done inclusively for the years with the year passed as an extra input variable. For ggF, the backgrounds were derived after the X_{Wt} cut, while the VBF background was derived after the X_{Wt} cut.²

Variable description	ggF	VBF
$\log(\Delta R_1)$: between the closest two HC jets	✓	
$\log(\Delta R_2)$ between the other two HC jets	✓	
$\log(p_T)$ of the 4th leading HC jet	✓	
$\log(p_T)$ of the 2nd leading HC jet	✓	
$\langle \eta \rangle$: average absolute value of the HC jets η	✓	✓
Number of jets in the event	✓	
$\log(p_{T,HH})$	✓	
ΔR_{HH}	✓	
$\Delta\phi$ between the jets in the leading HC	✓	
$\Delta\phi$ between the jets in the subleading HC	✓	
$\log(X_{Wt})$	✓	✓
Trigger bucket index	✓	✓
Year index		✓
Second smallest ΔR between the jets in the leading HC (out of the three possible pairings)		✓
Maximum di-jet mass out of the possible pairings of HC jets		✓
Minimum di-jet mass out of the possible pairings of HC jets		✓
Energy of the leading HC		✓
Energy of the subleading HC		✓

Table 11.1: Set of input variables used for the $2b$ to $4b$ reweighting for the ggF and VBF channels. The variables included in the background estimate are denoted with a checkmark.

Figure 11.3 shows some features comparing $2b$ and $4b$ distributions before the reweighting, and also evaluating the reweighting’s prediction for the 2018 background model with the ggF selection. The NNs shown here are both trained and evaluated in CR1, so these metrics evaluate how well the NNs have optimized in the training region, and don’t yet assess how well it will extrapolate into the SR. There is good modelling both for the input variables ΔR_{jj} Close and ΔR_{jj} Not Close, and also

²We saw for ggF that we had the same performance whether we trained before or after the X_{Wt} cut.

for m_{HH} , which was not one of the input reweighting variables. This gives us confidence that the selected variable set is sufficient for capturing the relevant event kinematics.

Figure 11.4 shows the performance of the VBF channel reweighting for a couple of the reweighting input variables, X_{Wt} and the energy of the leading HC, as well as the m_{HH} discriminating variable. The VBF channel is more statistically limited, and the impact of the reweighting correcting the m_{HH}

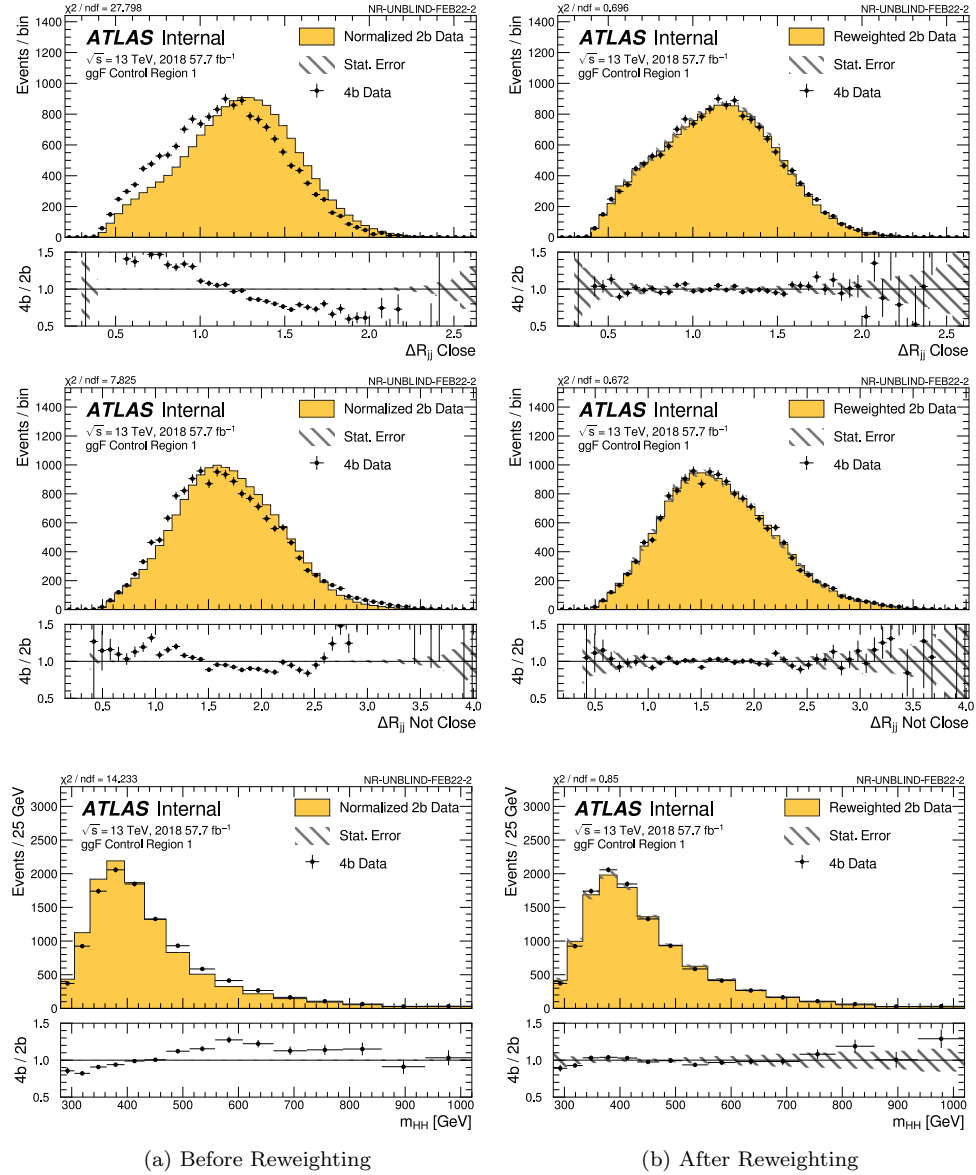


Figure 11.3: Distributions of ΔR between the closest Higgs Candidate jets, ΔR between the other two (training variables) and the invariant mass of the di-Higgs system (non-training variable) before (left column) and after (right column) the CR 1 derived reweighting for the 2018 Control Region 1.

distribution is not as dramatic as we see for the ggF channel, although the corresponding non-closure is covered by the corresponding errors on the precision of these reweighting maps (which will be discussed in Chapter 11.3.1).

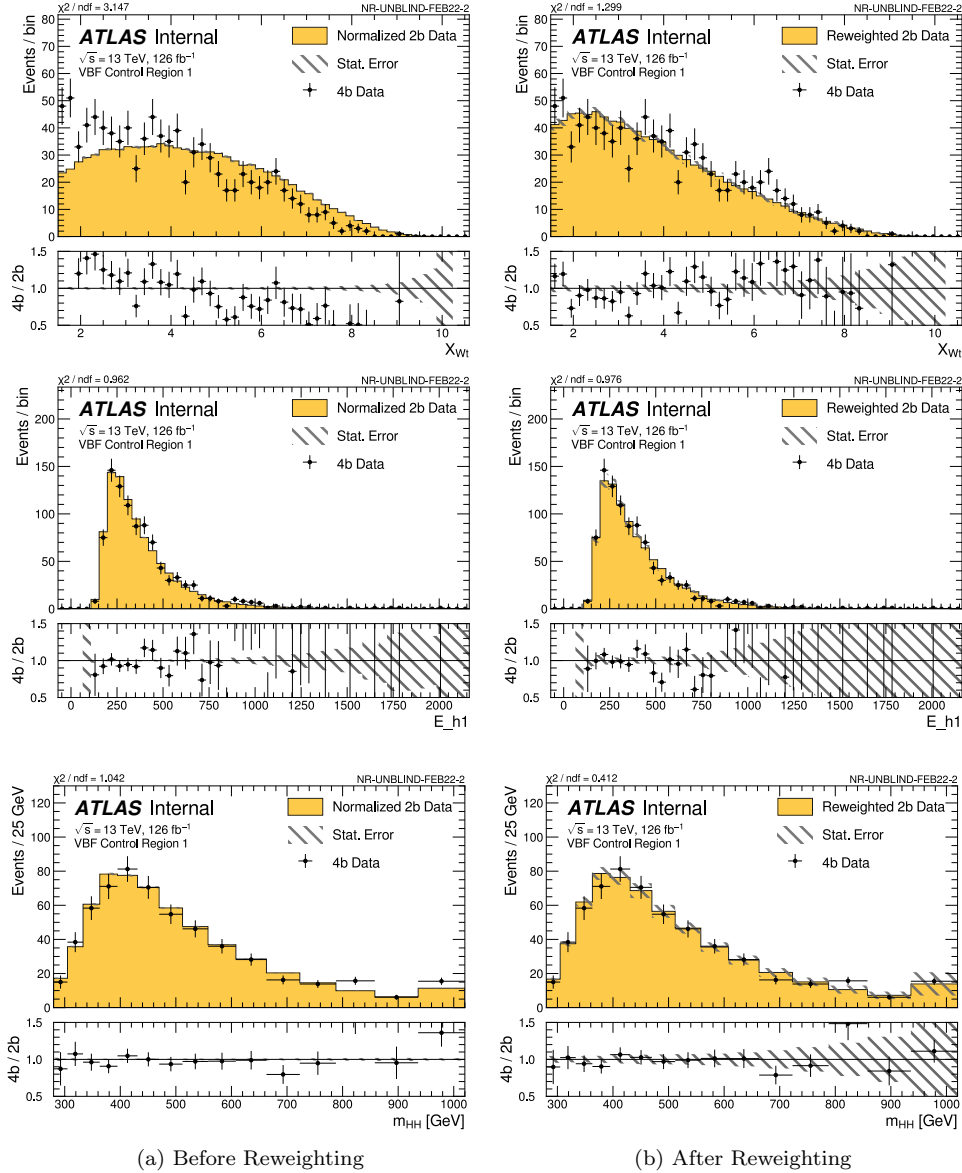


Figure 11.4: Distributions of the top veto variable, X_{Wt} , the energy of the leading Higgs candidate (training variables), and the invariant mass of the di-Higgs system (non-training variable) before (left column) and after (right column) CR 1 derived reweighting for the VBF Control Region 1.

Additional categorization was one of the key improvements from this analysis iteration, with $\Delta\eta_{HH}$ and X_{HH} categories for ggF and $\Delta\eta_{HH}$ categories for VBF. We checked the performance of the reweighting in CR1 for the m_{HH} distribution in each category. Both of these use a reweighting derived inclusively and then apply it to all of the categories. The X_{HH} categories are $[0, 0.95, 1.6]$, but $X_{HH} > 1.6$ in CR1 (since $X_{HH} < 1.6$ defines the SR). So for the ggF background estimate in CR1, we checked the $\Delta\eta_{HH}$ categorization for each of the years in Figure 11.6. The modeling with the VBF categorization is shown in Figure 11.5. Both of the channels have good closure in the CR1 training region.

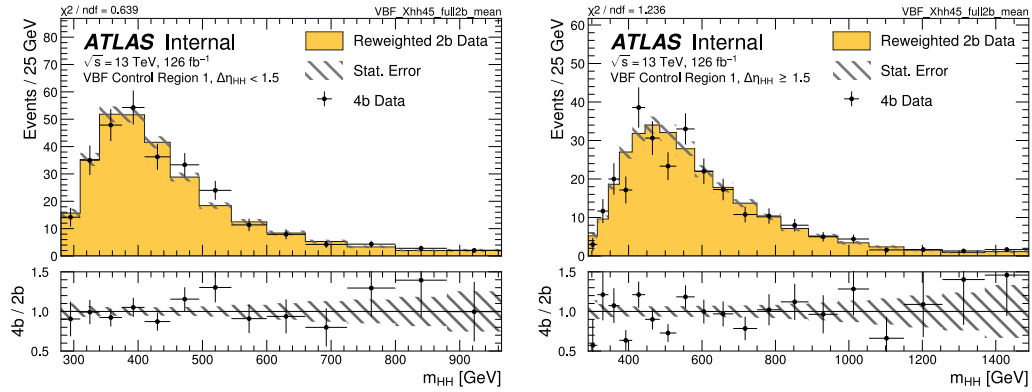


Figure 11.5: Distributions in CR1 after the reweighting and $\Delta\eta_{HH}$ categorization for VBF. Bootstrap and Poisson errors are included.

11.2 Background Composition

11.2.1 $t\bar{t}$

Although this background estimate is derived inclusively, $t\bar{t}$ is $\approx 10\%$ of our anticipated background, with proportions relative shown in Table 11.2. The relatively small contribution justifies our inclusive treatment of the QCD and $t\bar{t}$ in the reweighting definitions, although as part of this thesis work Appendix F shows how we could modify reweighting to account for these components separately in future iterations as we gain more sensitivity to the SM HH signal.

	ggF selection			
	2016	2017	2018	VBF
pre- X_{Wt}	19.5%	17.6%	18.2%	12.2%
post- X_{Wt}	10.4%	9.5%	10.4%	8.1%

Table 11.2: Percentage of the data-driven background estimate expected for $t\bar{t}$ events.

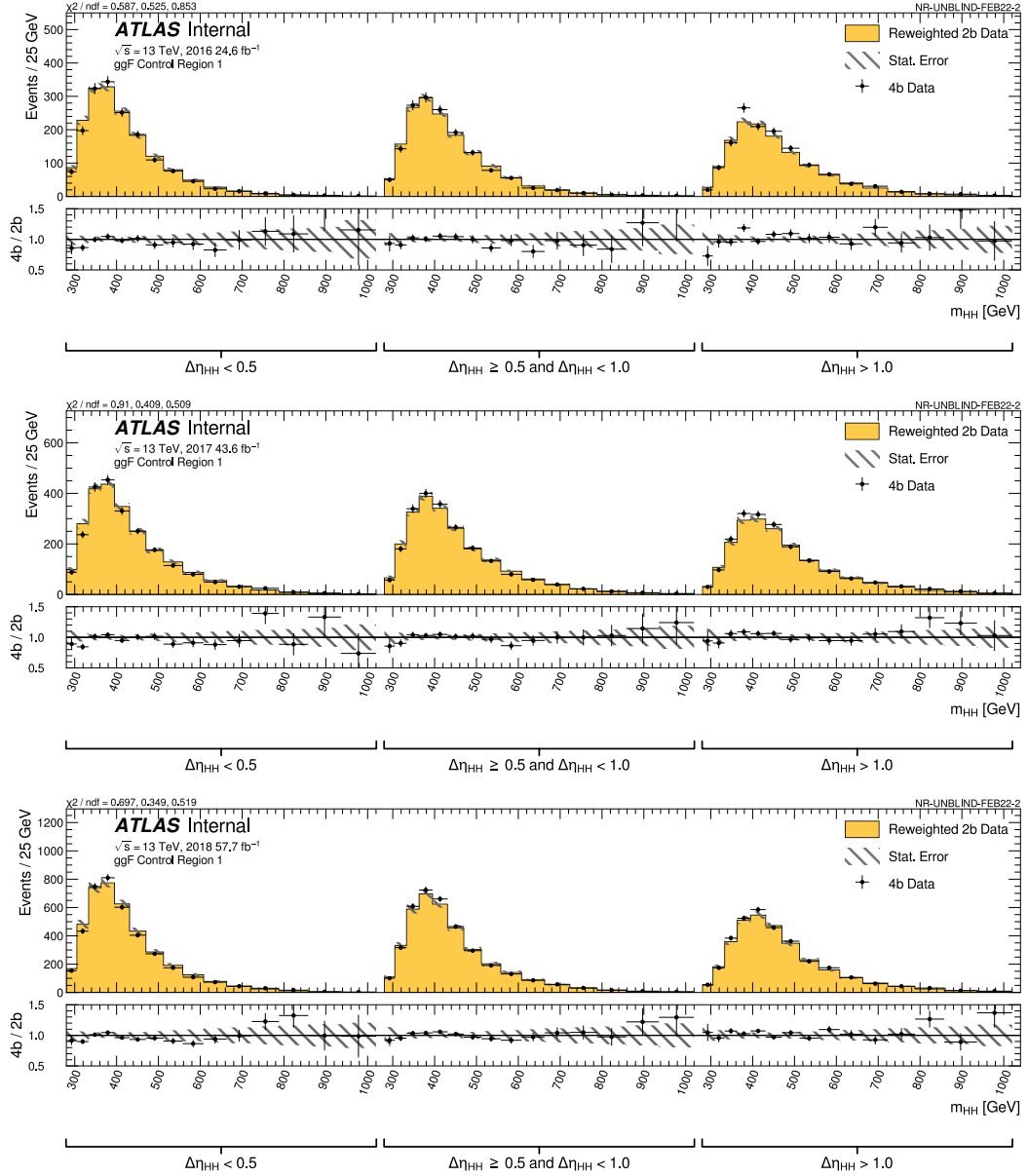


Figure 11.6: Distributions in CR1 after reweighting and $\Delta\eta_{HH}$ categorization for each year of the ggF background estimate, 2016 (top), 2017 (middle), and 2018 (bottom). Bootstrap and Poisson errors are included.

11.2.2 QCD flavor composition

Although we didn't trust the precision of the Pythia QCD for defining the background model, it can still give us a first-order estimate of the heavy flavor quark compositions in the 4b signal region. Figure 11.7(a) shows the m_{HH} QCD distribution, with the separate components showing the origin of the four b -tagged jets that reconstruct the HH system. The yellow template shows that 70% of the QCD background actually has 4b-quarks. When the analysis jets do not come from four truth b -jets, 22% of the time a c -jet was mistagged as a b -jet, and the remaining 8% of the time a light jet was mistagged as a b jet.

Figure 11.7(b) shows these same truth origin categories, but with each m_{HH} bin normalized to unity. We see an increase in the c -jets for $m_{HH} < 350$ GeV, since Figure 6.22(b) shows that the c -jet mistag rate for the DL1r tagger increases as the c -jet p_T decreases. For higher m_{HH} , we get a higher proportion from the mistagged light jets, which is understood as Figure 6.24(c) shows that the light jets mistag rate increases with jet p_T .

These investigations show how improving the b -taggers will help the next iterations of this analysis. As 30% of our background comes from mis-tagged jets, and reducing the c -jets is most important for decreasing the “fake b -jets” background yield.

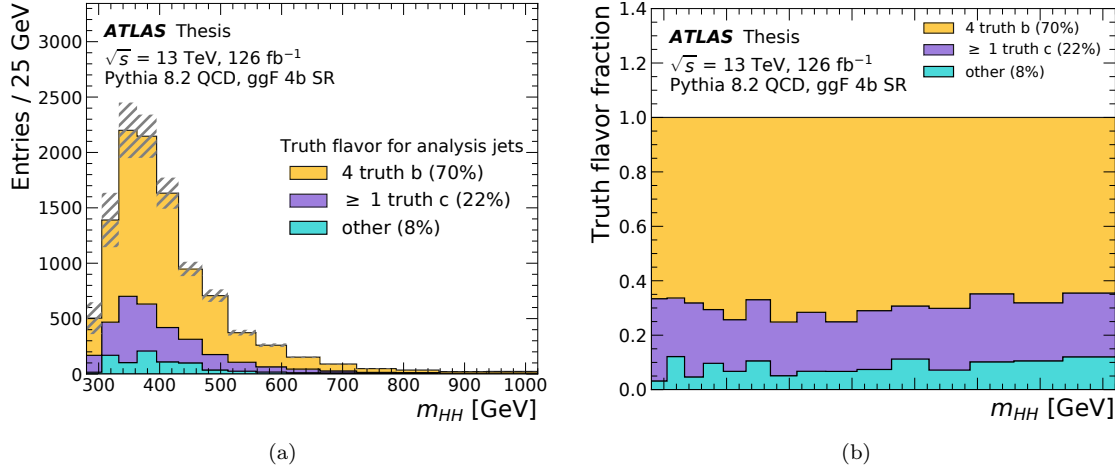


Figure 11.7: The flavor composition from Pythia QCD of the leading four p_T b -jets used to reconstruct the HH system (“analysis jets”) in the ggF 4b SR, as a function of m_{HH} . The left plots shows the m_{HH} distribution while the right plot normalizes each m_{HH} bin to unity.

11.3 Background systematics

This section describes the custom background modeling uncertainties we assess:

- stochasticity in the NN optimization (Chapter 11.3.1),
- choice of CRs for deriving these reweightings (Chapter 11.3.2),
- and (for ggF-only) a remaining extrapolation uncertainty from reweighting into another (lower) b -tagged target distribution (Chapter 11.3.3).

Chapter 11.3.4 compares the relative size of each of these contributions in the background model.

11.3.1 Deep ensembles + bootstrap error

The loss landscape of NNs is non-convex with many local minima. Different initializations for the NN weights will converge to different minima [161], as visualized in Figure 11.8. Averaging an ensemble of randomly initialized NNs gives a more robust prediction for the nominal estimate, and the variance over these predictions' defines an uncertainty on the function, called the *deep ensembles error* [127].

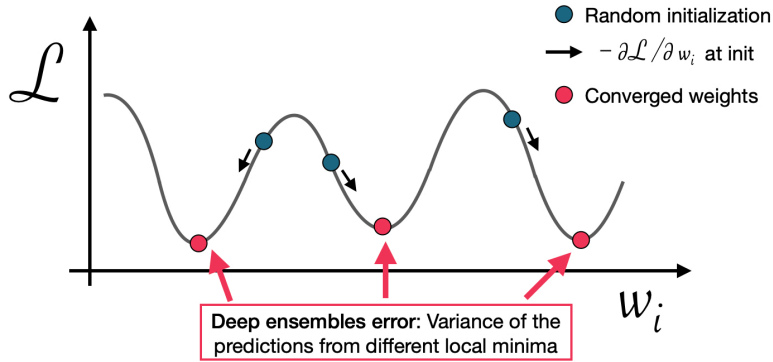


Figure 11.8: One dimensional non-convex, multi-modal loss function. The turquoise dots are randomly initialized weights w_i , and the black arrows are the gradients updates that minimize the loss. Each of the local minima (in the pink dots) is a different parametrized function, and the variance of these functions' predictions is the deep ensembles error. Modified from [161].

We also assess an error bar due to the finite training statistics. For a training dataset with N events, sampling this dataset N times with replacement defines a *bootstrap sample*. Since for $4b$ the training datasets are large ($\mathcal{O}(10^6)$ events), this sampling procedure can be simplified

$$\lim_{N \rightarrow \infty} \text{Binomial}(N, \frac{1}{N}) = \text{Poisson}(N \cdot \frac{1}{N}) = \text{Poisson}(1),$$

so each bootstrap samples a weight for each event from $w \sim \text{Poisson}(1)$. Each NN is trained on a separate bootstrap, and the variation in the NN predictions from these bootstraps is the *bootstrapping uncertainty*.

For the $4b$ background estimate, each model is trained 100 times, and the nominal estimate is the average of these predicted weights. The deep ensembles uncertainty dominates compared to the bootstrapping uncertainty, but in the text and figures, “bootstrap” refers to the error from the combination of these two sources. Figure 11.9 shows the predicted m_{HH} distribution for each of these 100 NN trainings, relative to the nominal estimate. Let h_i refer to the histogram for a single NN training, and \hat{h} denote the average of these histograms. In each bin, the standard deviation of the yields from the 100 h_i histograms defines the deep ensembles + bootstrap error bar. I.e, in bin j , the error is:

$$\sigma_j = \frac{1}{100} \sum_{i=1}^{100} \left(h_{ij} - \hat{h}_j \right)^2.$$

This σ/\hat{h} is shown in the pink line on Figure 11.9. The σ_j is then summed in quadrature with the 2b statistical error for each bin,³ and this combined uncertainty is encoded in the profile likelihood fit with a per-bin Poisson nuisance parameter.

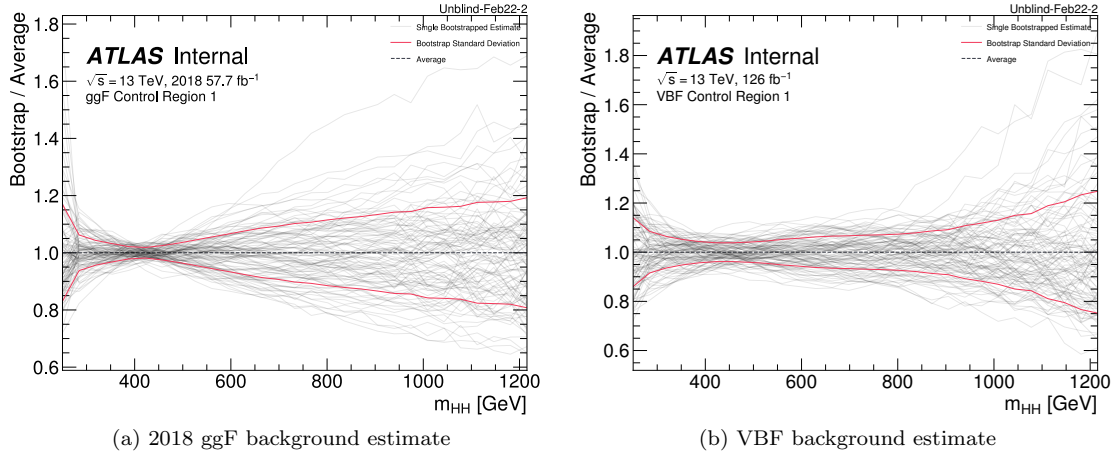


Figure 11.9: Illustration of the deep ensembles error, shown as a relative error with respect to the nominal estimate. Each grey line is from the m_{HH} prediction for a single NN training, and the solid pink line shows the standard deviation of the gray histograms.

³ Although in practice these bootstrap trainings are correlated between the bins, these are ignored as including the correlations did not have a significant impact on our results.

11.3.2 Choice of control region

As introduced in Chapter 11.1, to account for the systematic bias associated with deriving the reweighting function in CR1, an alternative background model is derived in CR2. The CR2-derived model uses the same prescription of training 100 NNs and using the average of the 100 trainings for the nominal estimate.

This yield in each bin of the CR2 estimate is additionally symmetrized:

$$\frac{\hat{h}_j^{CR2,sym}}{\hat{h}_j^{CR1}} = 1 - \left(\frac{\hat{h}_j^{CR2}}{\hat{h}_j^{CR1}} - 1 \right) = 2 - \frac{\hat{h}_j^{CR2}}{\hat{h}_j^{CR1}} \quad \Rightarrow \quad \hat{h}_j^{CR2,sym} = 2\hat{h}_j^{CR1} - \hat{h}_j^{CR2} \quad (11.4)$$

where \hat{h}_j^{CR1} is histogram of the average of the CR1 networks (the “nominal estimate”), \hat{h}_j^{CR2} is histogram of the average of the CR2 networks (the “alternative estimate”), and $\hat{h}_j^{CR2,sym}$ is the symmetrized for the ratio of about unity. This prescription for defining these CR12 error templates is summarized in Figure 11.10. We have four background estimates (three for each year of the ggF selection, and one for the VBF selection), and this extrapolation uncertainty is derived separately for each of these reweightings. We also define the templates separately for each of the $\Delta\eta_{HH} + X_{hh}$ categories with the ggF selection, and the $\Delta\eta_{HH}$ categories for the VBF selection.

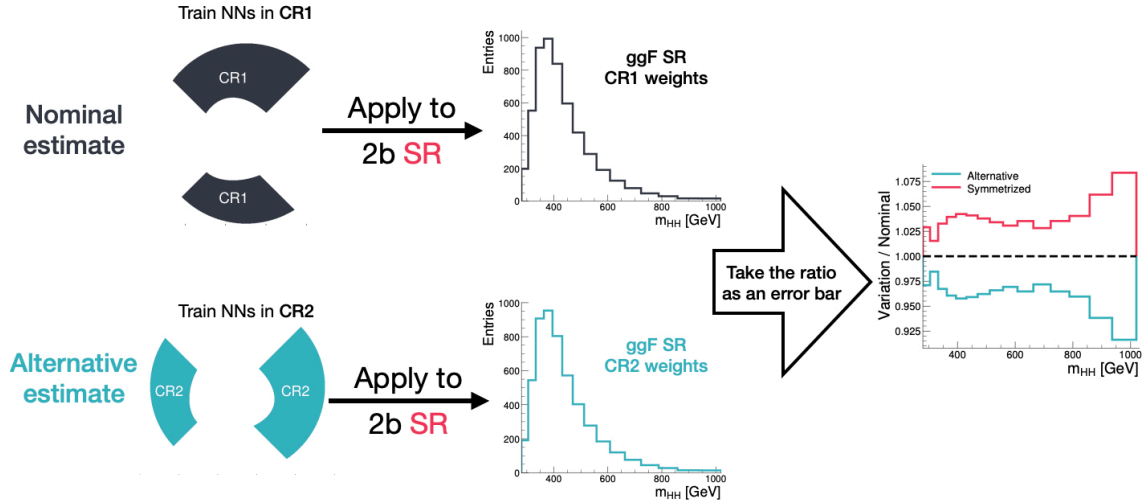


Figure 11.10: Illustration of the derivation of the shape systematic. The top row shows the derivation of the nominal estimate from the CR1 regions, and the bottom row shows the alternative estimate derivation from the CR1 region. The ratio between these predictions is the “CR12 shape systematic”. The symmetrized alternative model is shown in the pink variation histogram on the right plot. These templates are defined for each of the $\Delta\eta_{HH}$ (+ X_{HH} for ggF) categories.

SR Nuisance Parameter decomposition

The right plot in Figure 11.10 shows that the uncertainties become larger for at m_{HH} and are smaller in the moderate m_{HH} regime from 500 – 800 GeV that drives our analysis sensitivity. Since different physics processes are relevant in these two regions – we allow the fit additional flexibility to account for different contributions for the CR1 versus CR2 variations. We account for these separate contributions by cutting on an event variable. In past iterations of the 4b analysis, $H_T = 300$ was used to factorize the low and high m_{HH} variations [116, 117].

For this analysis, the separate contributions are considered by dividing the 4b SR into four quadrants, shown in Figure 11.11. The four colors indicate the four separate quadrants in the North, West, South and East directions. A variation template is constructed for each of these quadrants. For the Q_N template, the alternative histogram applies the CR2 weights for events in Q_N , and then applying the CR1 weights for events in Q_S , Q_E and Q_W . The templates for the other quadrants are defined similarly with the CR2 weights applied to the relevant quadrant.

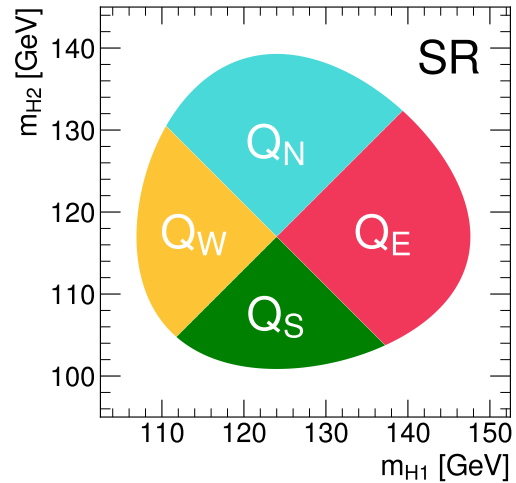


Figure 11.11: SR quadrants chosen to derive the four background variation NPs.

The ratio between the histograms from the nominal estimate and alternative histograms for each of these four quadrants is used to define an uncertainty from extrapolating from the CR into the SR, called the “CR12 extrapolation uncertainty”. The corresponding templates for each of for each of these quadrants with the 2018 ggF background estimate are shown in Figure 11.12. Each plot shows a separate SR quadrant, with the m_{HH} distribution with both the $\Delta\eta_{HH}$ and X_{HH} categorization. The histograms defined with nominal estimates are shown in black. Histograms using the CR2 estimate are shown in turquoise, and the symmetrized histogram is shown in pink. The subpanels show the ratio of these alternative histograms to the nominal one. The Q_N template (in the top row) captures most of the variation at low and moderate m_{HH} , while the Q_E template accounts for most of the variation in the high m_{HH} tails.

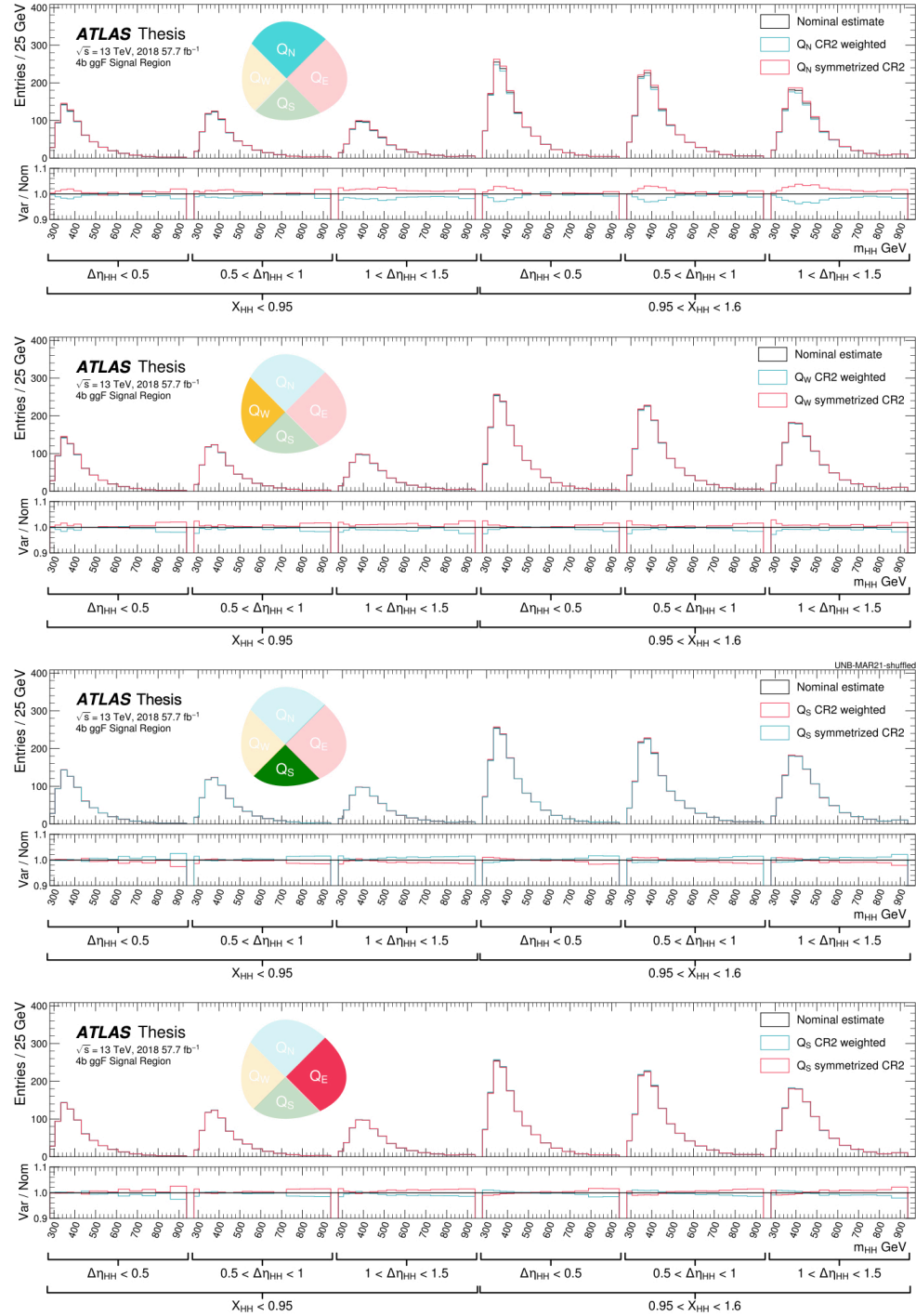


Figure 11.12: Templates for the CR12 extrapolation systematic for the 2018 ggF discriminant in the SR NP quadrants for Q_N , Q_W , Q_S , Q_E (ordered top to bottom). The nominal histogram is shown in navy, while the alternative histograms defining the $\pm 1\sigma$ variations are shown in the pink and turquoise histograms. The ratio panel shows the ratio of these variation histograms to the nominal.

Figure 11.13 overlays the templates for each of the $\Delta\eta_{HH}$ and X_{HH} kinematic categories. Each row depicts a different SR quadrant and each column denotes a separate year of the ggF background estimate. In each plot the colors show the $\Delta\eta_{HH}$ categories, while the solid and dashed lines indicate the X_{HH} categories. The variations across the $\Delta\eta_{HH}$ X_{HH} categories have similar variations for the m_{HH} templates. Thus, in the profile likelihood fit, we correlate the NPs across the kinematic categories.

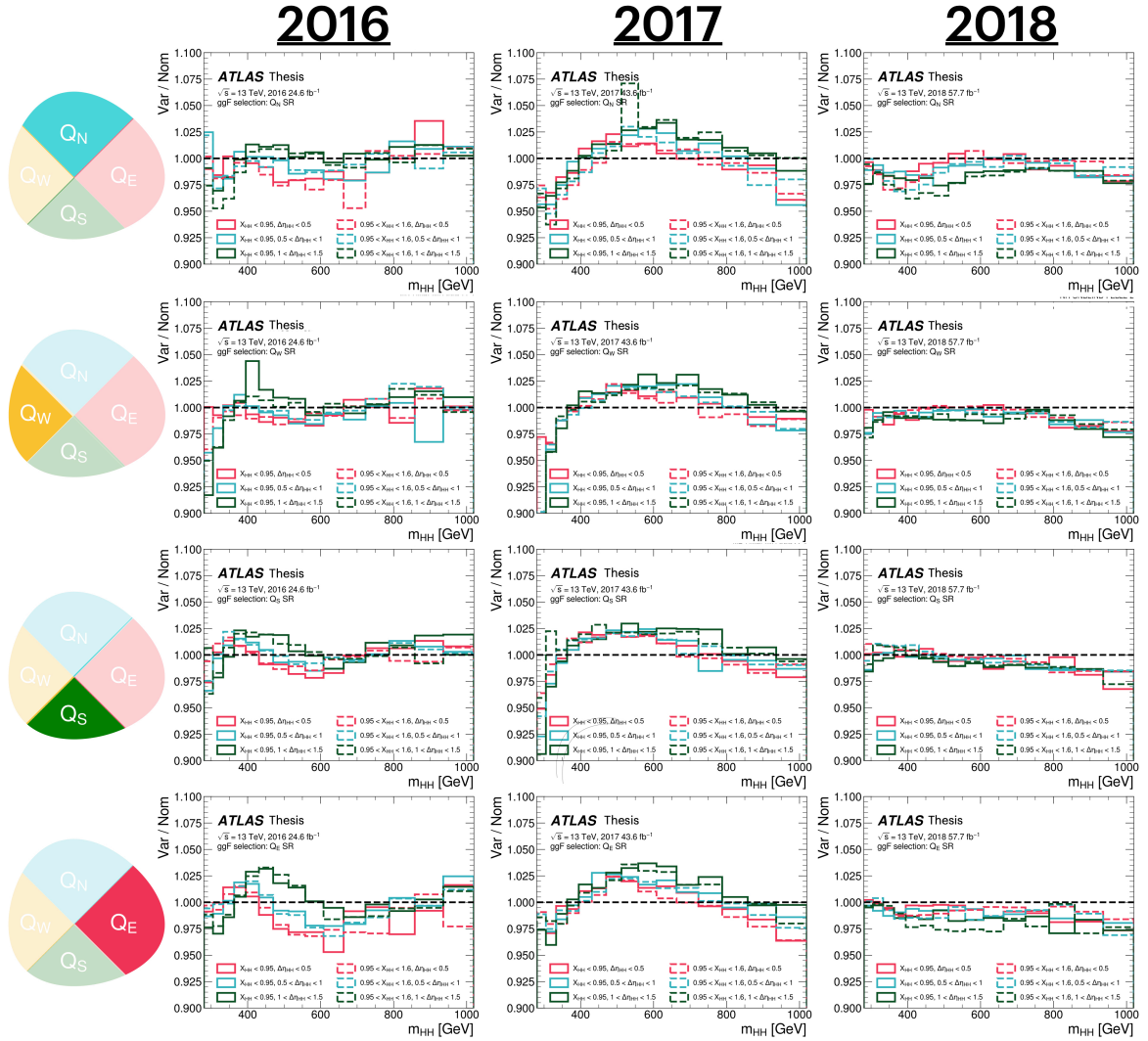


Figure 11.13: Impact of background shape nuisance parameter variation on m_{HH} in different kinematic categories for the ggF channel. Each column is a different year for the ggF channel templates while the rows show the SR NP quadrants.

11.3.3 3b1f non-closure uncertainty

The 3b1f (3b + 1 fail) region has three jets b -tagged at the 77% working point, and all other jets failing the (loosest) 85% b -tagging working point. The the highest p_T non- b -tagged jet is used as the 4th analysis jet. The background estimate procedure was repeated in this region, reweighting 2b events into 3b1f, and we saw a non-closure in the ggF channels that was not covered by the bootstrap+deep ensembles and the CR12 extrapolation systematic uncertainties described in the previous two sections. The statistical errors are larger VBF channel, and there was no non-closure observed in the 3b1f VBF SR region. We included an uncertainty to cover the non-closure in the 3b1f ggF region, which is defined as follows:

- In each analysis category, the ratio of the 3b1f data and the reweighted 2b m_{HH} histograms was taken.
- If the deviation from unity is larger than the quadratic sum of the deep ensembles, bootstrap, 2d statistical error, the CR12 extrapolation systematic uncertainties, and the 3b1f Poisson uncertainty, the bin-by-bin residual is taken as an uncertainty template.
- This template is smoothed by averaging each bin with its two neighboring bins and defines the “3b1f non-closure uncertainty” (as will be shown in the red lines on Figure 11.14).

11.3.4 Summary of Background Modeling Uncertainties

The contribution of each uncertainties on the yield of background estimates are shown in Table 11.3.

4b SR	Counts				Relative [%]			
	2016	ggF channel 2017	2018	VBF	2016	ggF channel 2017	2018	VBF
Bkg Yield	3211.1	4492.7	7720.0	354.6	–	–	–	–
Total Stat Error	225.8	308.9	483.0	32.3	7.03	6.88	6.26	9.10
Shape Q_N	–25.3	–24.7	–141.3	–1.2	–0.79	–0.55	–1.83	–0.35
Shape Q_W	–12.1	–4.8	–55.8	–0.8	–0.38	–0.11	–0.72	–0.23
Shape Q_S	16.2	43.0	7.0	–0.5	0.50	0.96	0.09	–0.15
Shape Q_E	7.0	14.8	–87.1	–0.3	0.22	0.33	–1.13	–0.09
3b1f NC	6.8	17.6	42.6	–	0.21	0.39	0.55	–
Bkg Syst Error	33.9	54.9	180.4	1.6	1.05	1.22	2.34	0.45
Total Bkg Error	228.3	313.7	515.6	32.3 ‘	7.11	6.98	6.68	9.11

Table 11.3: Magnitude of error components for the ggF and VBF channels in the Signal Region. Total statistic error is the quadrature sum of bootstrap and 2b stat errors. The errors are just showing the impact on the overall normalization without the categorization.

The impact of each of these uncertainty sources on the shape of the discriminant are shown in Figure 11.14 for the ggF channels and Figure 11.15 for the VBF channels.

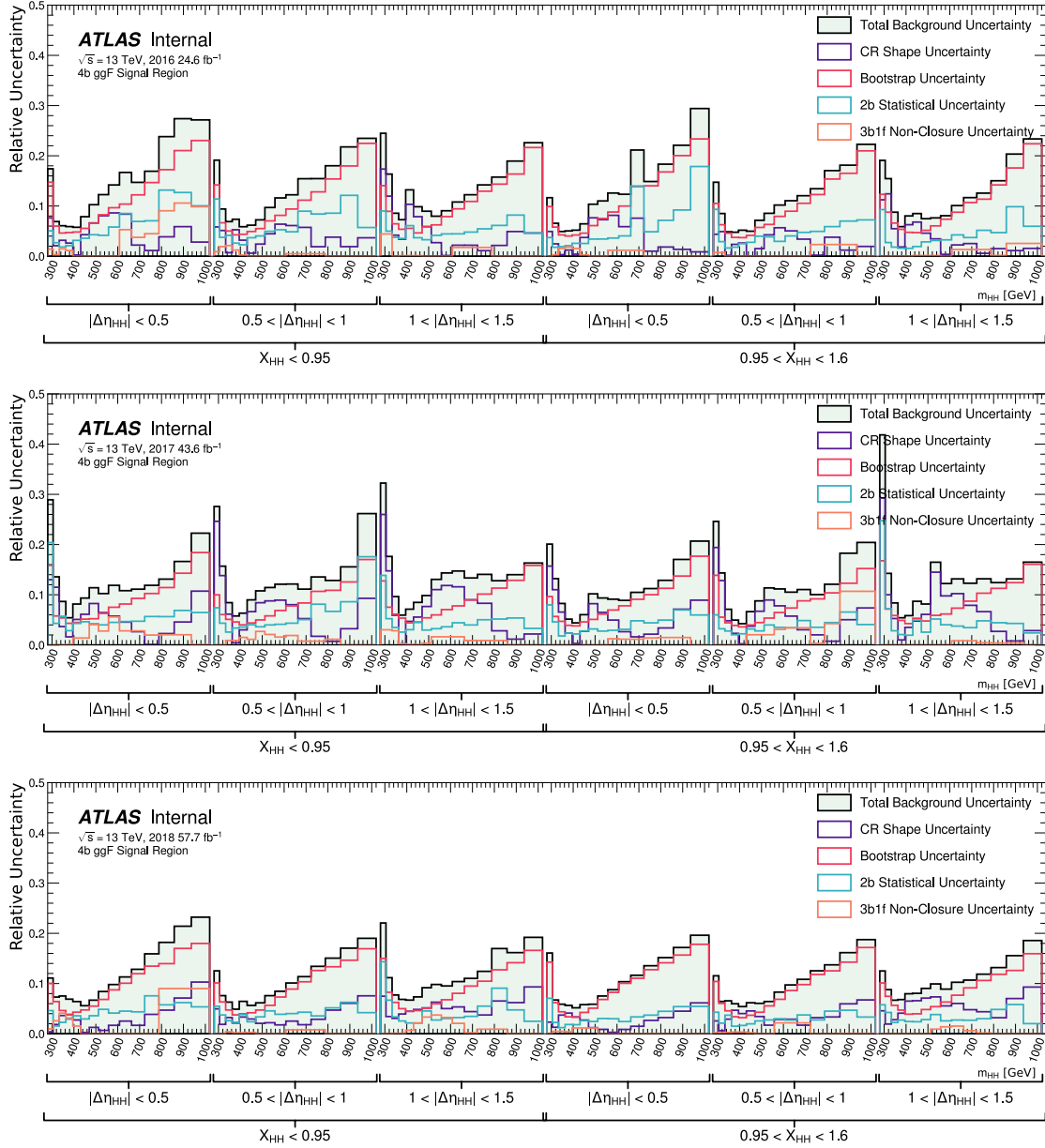


Figure 11.14: Relative error contributions of the background for the ggF discriminant for the 2016 (top), 2017 (middle) and 2018 (bottom) background estimates.

The statistical error dominates for high m_{HH} while the shape systematic from the difference in the CR1 and CR2 estimates contributes more in the moderate m_{HH} region that drives our analysis sensitivity. Although the stat error (from the limited 2b data statistics) is negligible relative to the bootstrap error in the bulk of the distribution, it becomes relevant in the high m_{HH} tail.

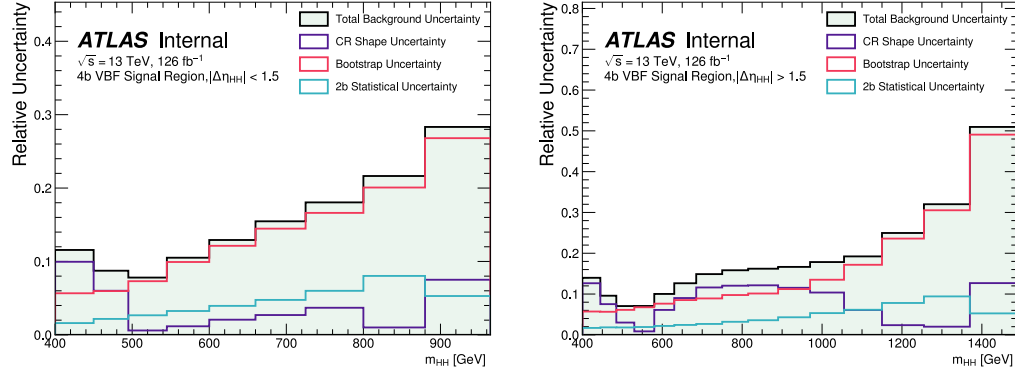


Figure 11.15: Relative error contributions of the background for the VBF discriminant.

Table 11.4 summarizes the categories that we have for the ggF and VBF analyses, and how this translates to the number of background NPs.

Categories	ggF channel	VBF channel
	$3 \text{ years} \times 3 \Delta\eta_{HH} \times 2 X_{HH}$ = 18 categories	$2 \Delta\eta_{HH} $ categories
Deep ensembles, bootstrap, and 2b stat nuisance parameters	# of bins	# of bins
CR12 extrapolation nuisance parameters	$3 \text{ years} \times (Q_N, Q_E, Q_S, Q_W) = 12$	$(Q_N, Q_E, Q_S, Q_W) 4$
3b1f non-closure nuisance parameters	18 (1 per category)	0

Table 11.4: Summary of the categorization strategy and background-related nuisance parameters.

11.4 Background validation

Since background estimation is the main challenge of this analysis, careful checks were done to validate the procedure. In this section, we focus on the validation results for the ggF channel. In Chapter 11.4.1 the $\Delta\eta_{HH}$ ggF cut is inverted, and in Chapter 11.4.2 the SR is shifted around the massplane to create validation regions. Chapter 11.4.3 concludes with evaluations in MC samples.

Since each of these validation regions have negligible signals, we could compare the validation’s “SR” to the background estimate to test the background systematic prescription defined in Chapter 11.3. For each of these test, the 100 NNs are retrained in both CR1 and CR2, and the background uncertainties we show include the bootstrap and deep ensembles and the CR12 extrapolation systematic. The 3b1f non-closure systematic is not included here as these validation tests had satisfactory closure without it.

11.4.1 Reversed $|\Delta\eta_{HH}|$

Since the ggF selection imposes a $\Delta\eta_{HH} < 1.5$ to suppress the QCD background, inverting this cut gives a signal-depleted region. The corresponding 4b 2018 m_{HH} and $\Delta\eta_{HH}$ marginals are shown in Figure 11.16, and shows good modeling within the background errors.

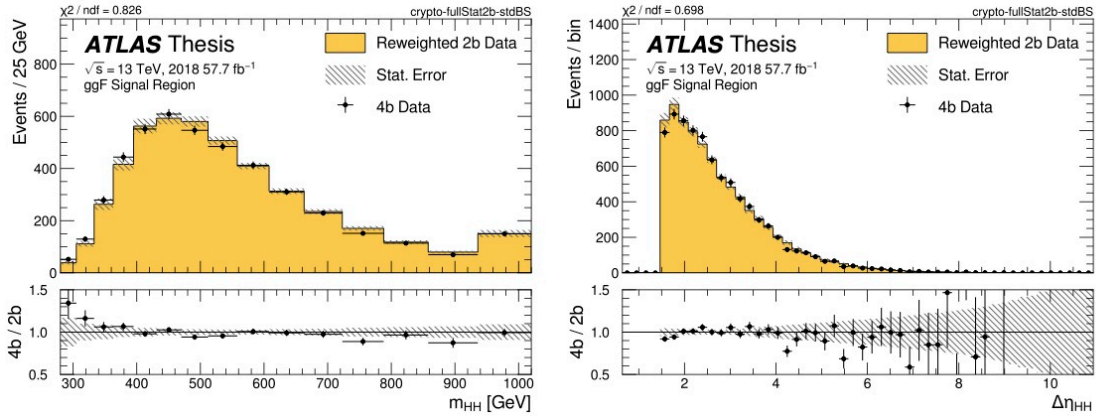


Figure 11.16: m_{HH} and $\Delta\eta_{HH}$ distributions for the rev $\Delta\eta_{HH}$ 2018 ggF data sample.

To also validate the analysis with the ggF categories, we choose $\Delta\eta_{HH}$ boundaries to keep equal background statistics between each of the regions:

- $1.5 < |\Delta\eta_{HH}| < 2.5$
- $2.5 < |\Delta\eta_{HH}| < 3.6$
- $|\Delta\eta_{HH}| > 3.6$

The corresponding high-level discriminants for each of the years is shown in Figure 11.17. The 4b data $\Delta\eta_{HH}$ is consistent with the reweighted 2b background within the background uncertainties.

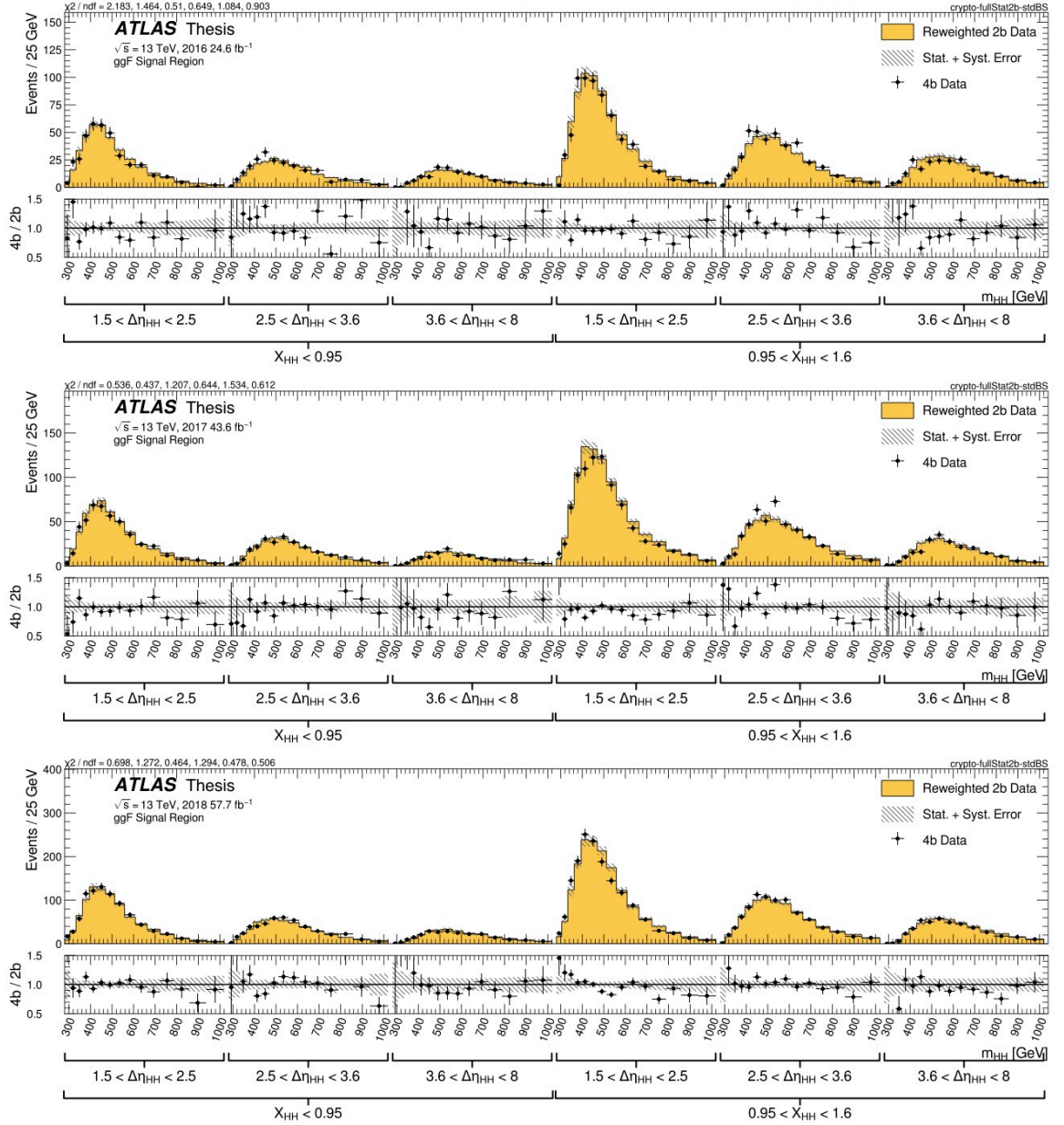


Figure 11.17: The ggF 4b rev $\Delta\eta_{HH}$ discriminants for the 2016 (top), 2017 (middle) and 2018 (bottom) background estimates.

11.4.2 Shifted regions

The signal region (and associated CRs) was also modified to move it around the massplane, as shown in Figure 11.18. The equation for the SR defining variable $X_{HH} = \sqrt{\left(\frac{m_{H1} - 124 \text{ GeV}}{0.1 m_{H1}}\right)^2 + \left(\frac{m_{H2} - 117 \text{ GeV}}{0.1 m_{H2}}\right)^2}$ has a radius that depends on m_{H1} , m_{H2} . For this study the HC resolutions in the X_{HH} formula were modified so that the SR size stays compatible with the standard SR for shifts around the massplane:

$$X_{hh,\text{shift}} = \sqrt{\left(\frac{m_{H1} - m_{H1,\text{center}}}{\sigma_{m_{H1}} m_{H1}}\right)^2 + \left(\frac{m_{H2} - m_{H2,\text{center}}}{\sigma_{m_{H2}} m_{H2}}\right)^2} \quad (11.5)$$

with $\sigma_{m_{H1}} = 0.1 \times \frac{124}{m_{H1,\text{center}}}$, $\sigma_{m_{H2}} = 0.1 \times \frac{117}{m_{H2,\text{center}}}$.

The CR boundaries shifts with the SR center, and the reweighting quadrants are defined analogously.

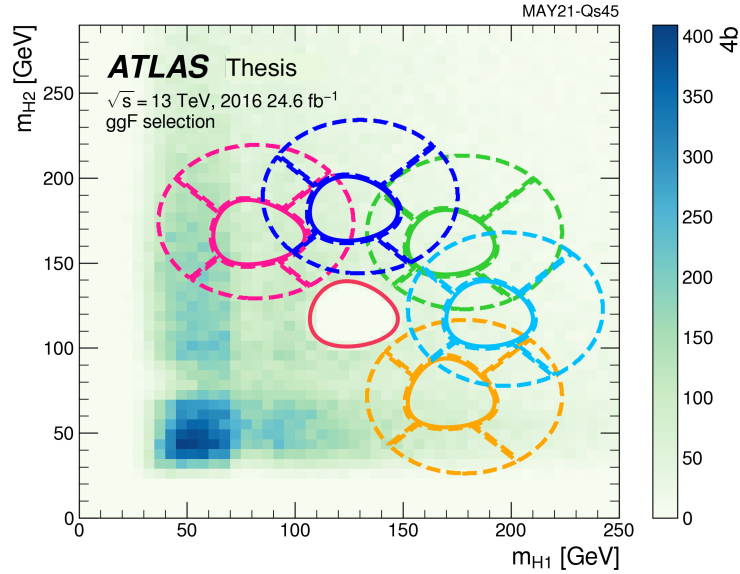


Figure 11.18: The shifted regions for the background validation, with the pink solid curve in the center showing the standard SR. The SRs are shown in the solid lines, and the corresponding control for deriving the background estimate and CR12 extrapolation error are shown in the dashed lines.

For the lower right SR, the lower quadrant for CR1 has a couple of issues. It (1) intersects with the W-veto from the X_{Wt} cut and (2) overlaps with the kinematic threshold because there are essentially no m_{H2} values below 25 GeV to populate the full lower quadrant. This made the extrapolating from the lower right CR1 quadrants into the SR more challenging than the extrapolation for the nominal SR. Since the CR2 (left, right) quadrants did not suffer from these two issues, in the following the CR2-derived weights are shown as the nominal set. Table 11.5 enumerates the shifted SR centers, and which CR is considered as the nominal.

(Shifted) Signal Region	m_{H1} center [GeV]	m_{H2} center [GeV]	nominal rw set
upper left	78	166	CR1
upper center	124	180	CR1
upper right	170	166	CR1
center right	188	117	CR1
lower right	170	68	CR2
standard	124	117	CR1

Table 11.5: Center locations for the shifted SRs validation study. The right-most column indicates which quadrants form the “nominal” background estimate.

The reweighting functions were derived for each of these regions and for all of the years, and Figure 11.19 shows the m_{HH} predictions compared to the observed 4b data for 2018. In the ratio panels, good agreement is seen with respect to the background uncertainties.

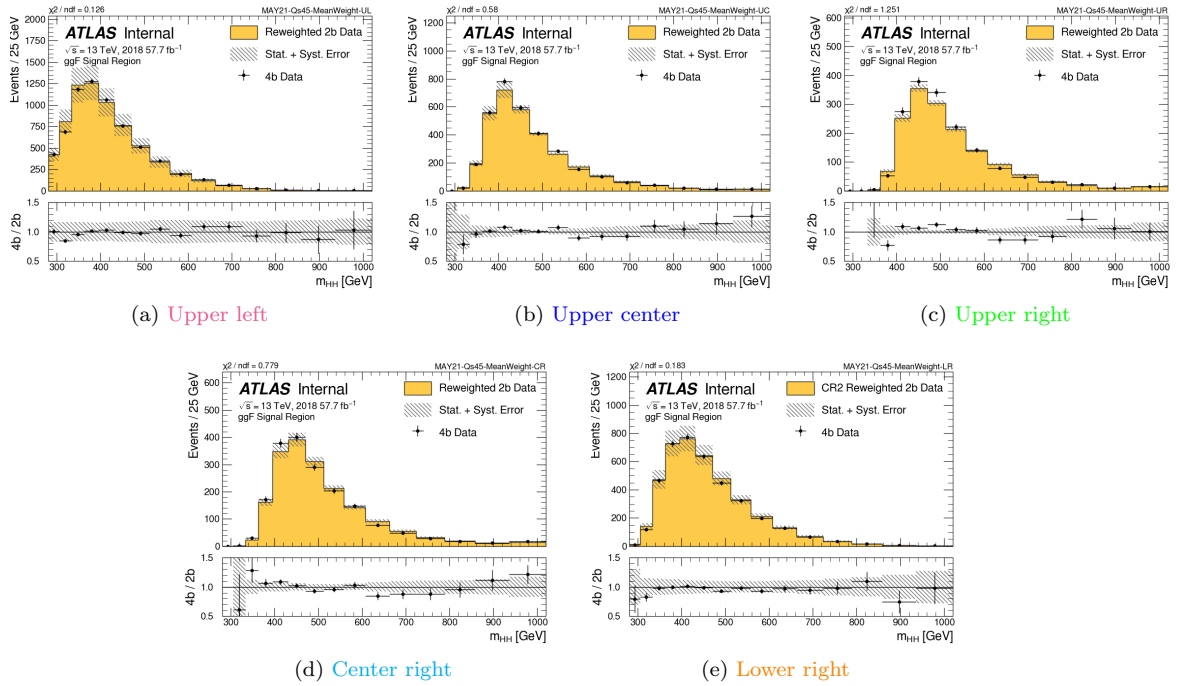


Figure 11.19: m_{HH} distributions of reweighted 2b data and 4b data in the shifted SRs for the 2018 background estimates. The background error bar includes the 2b Poisson, bootstrap+deep ensembles, and the CR1 / CR2 extrapolation systematics.

A comparison of the predicted and observed yields for all of the years are shown in Table 11.6.

Shifted Region	Year	4b Yield	Background Prediction	Deviation (μ_{norm})
upper left	2016	4068 ± 64	4101.6 ± 474.4	0.07
	2017	5586 ± 75	5843.4 ± 514.3	0.50
	2018	9421 ± 97	9559.7 ± 1051.2	0.13
upper center	2016	2197 ± 47	2086.3 ± 52.4	-1.57
	2017	3017 ± 55	2987.9 ± 72.8	-0.32
	2018	5161 ± 72	5058.6 ± 141.6	-0.65
upper right	2016	1182 ± 34	1125.8 ± 38.8	-1.08
	2017	1738 ± 42	1684.4 ± 40.2	-0.93
	2018	2831 ± 53	2732.7 ± 48.4	-1.37
center right	2016	1305 ± 36	1310.2 ± 43.6	0.09
	2017	1922 ± 44	1951.2 ± 73.7	0.34
	2018	3098 ± 56	3108.0 ± 98.6	0.09
lower right	2016	2658 ± 52	2664.2 ± 300.0	0.02
	2017	3635 ± 60	3814.4 ± 326.7	0.54
	2018	6084 ± 78	6241.1 ± 491.1	0.32

Table 11.6: 4b and background prediction in the signal region in the shifted regions in 2016. The error of background prediction includes the 2b poisson statistic error, the bootstrap error and the shape systematic error.

The agreement between the observation and the prediction is quantified by:

$$\mu_{\text{norm}} = \frac{(N_{\text{bkgd}} - N_{\text{target}})}{\sigma_{\text{stat}}} \quad (11.6)$$

where σ_{stat} includes the statistical errors from the 4b Poisson error (on the 4b data), the 2b Poisson error (on the background template), and the bootstrap and deep ensembles error – all summed in quadrature. Figure 11.20 shows the histogram of these μ_{norm} the across years and shifted regions. The Gaussian fit to the μ_{norm} s has a mean of -0.08 ± 0.17 , which is consistent with zero indicating no bias in the yields. The standard deviation of the μ_{norm} s is 1.19 ± 0.17 , which is consistent with zero and indicates that these errors accurately describe the underlying uncertainties.

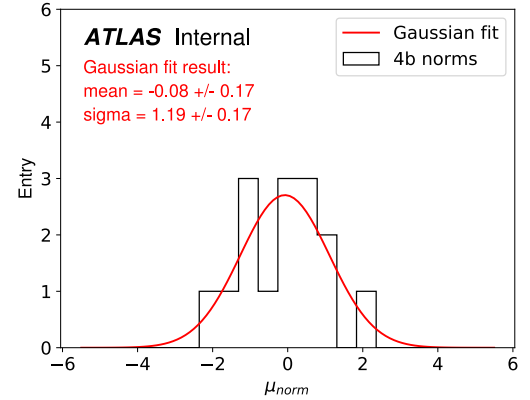


Figure 11.20: μ_{norm} distribution in the shifted regions. 4b normalizations (black) and the gaussian fit (red) are shown.

Finally, we tested the CR12 extrapolation systematics by taking each of these shifted SRs and finding the best fit for the nuisance parameters. This allowed us to test whether these best-fit values agreed with our expectations. In the *pull plot* in Figure 11.21, the blue bands correspond to the $\pm 1\sigma$ variations and the yellow bands correspond to the $\pm 2\sigma$ variations. A pull towards $+1\sigma$ indicates that the CR2-defined alternative model is a better description of the data than the CR1-defined nominal model, and large pulls indicate that we have underestimated our uncertainties.

In Figure 11.21 we fit the likelihood to the observed 4b shifted SRs data. The background nuisance parameters are allowed to vary within their constraints and the signal strength is fixed to $\mu = 0$. The nuisance parameters shown are for each of the years (16,17,18) and with the SR quadrants (E,N,S,W). Since the best fit values are within the 1σ error bands of the inputted templates, this gave us confidence moving that we have a good description of these uncertainties. This also bolstered our confidence in our SR quadrant nuisance parameter decomposition scheme.

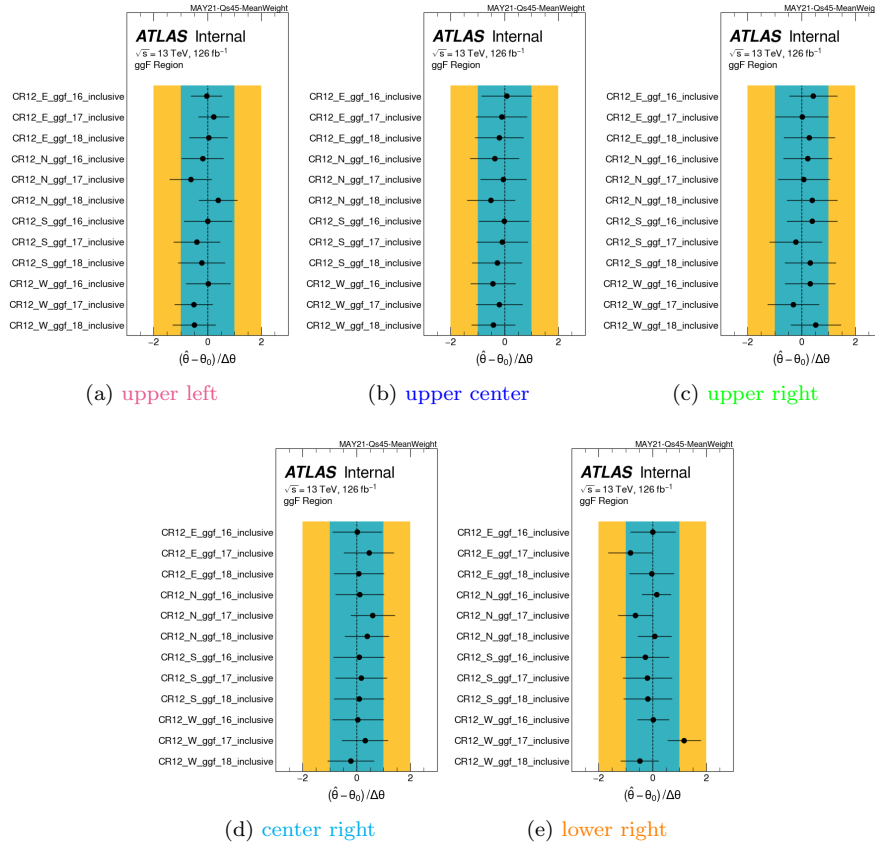


Figure 11.21: Pull plots for fits with background nuisance parameters floating for the five shifted regions. The signal strength is fixed to $\mu = 0$.

11.4.3 MC validation

Finally, we used the MC QCD and $t\bar{t}$ samples to devise two additional tests of the reweighting before unblinding. Although we didn't trust the MC as a background model, testing the reweighting on the MC sample allows us to test the SR extrapolation performance with the same cuts as the 4b SR.

Validation with data trained networks

The CR1 data trained NNs (a.k.a, the “nominal” background estimate) are evaluated on the MC 2b SR, and compared to the MC 4b SR. This test isn't perfect because of known discrepancies between data and MC, but this allows a test of the nominal NNs since the 4b SR MC is more similar to the 4b data than the other validation tests that use different kinematic cuts.⁴ Figure 11.22 shows m_{HH} validation plots. The MC samples in the bulk of the distribution have 1/10 of the events expected data (see Table 11.7), so some of the 4b errors are quite large, but overall the comparison seems reasonable.

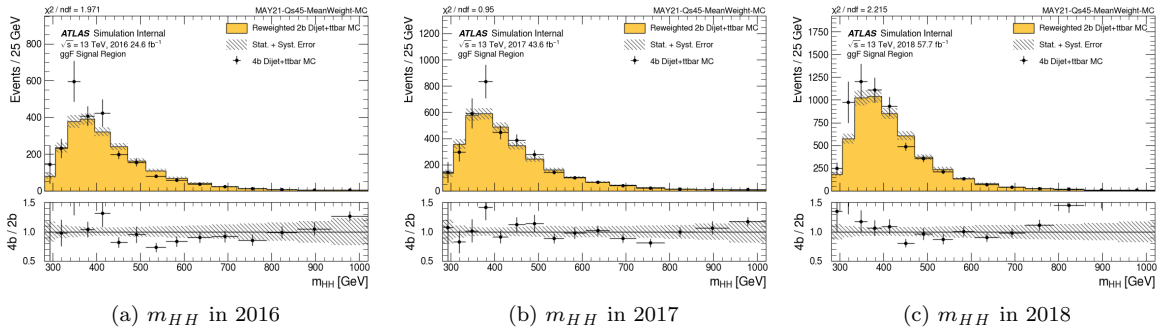


Figure 11.22: m_{HH} data reweighted events evaluated on the QCD and $t\bar{t}$ MC samples. The background error includes the deep ensembles, 2b stat, and CR1/CR2 systematic errors.

	4b CR1 events			2b CR1 events		
	2016	2017	2018	2016	2017	2018
Data	6,666	9,416	15,959	743,598	660,268	1,579,192
JZ2 (79.0 % of QCD)	752	922	1,445	82,723	61,484	150,628
JZ3 (20.4 % of QCD)	2,215	3,008	3,868	118,435	149,487	166,884
JZ4 (0.6 % of QCD)	2,142	3,315	1,755	43,480	71,732	57,301

Table 11.7: (Unweighted events) in CR1 for both data and the QCD samples. Indicated in the QCD sample name is the contribution to the (weighted) event yield. The highlighted rows indicate that the JZ2 MC sample has only 1/10 of the data statistics.

⁴Inverting the $\Delta\eta_{HH}$ cut makes the m_{HH} distribution harder, and the shifted SRs with one or both of the HC masses increasing also make the m_{HH} distribution larger. The MC 4b SR, on the other hand, uses the exact same kinematic cuts as the data 4b SR.

Validation with MC (re)trained networks

Here the background estimates were retrained for the MC QCD + $t\bar{t}$ dataset. Figure 11.22 shows m_{HH} distributions for evaluating these MC trained networks on the 2b SR MC, and comparing to the 4b SR MC prediction. The smaller training stats are reflected by the correspondingly larger error bar on the background estimate, but overall the data and background model are in reasonable agreement, and also compatible with the data trained estimate of Figure 11.22.

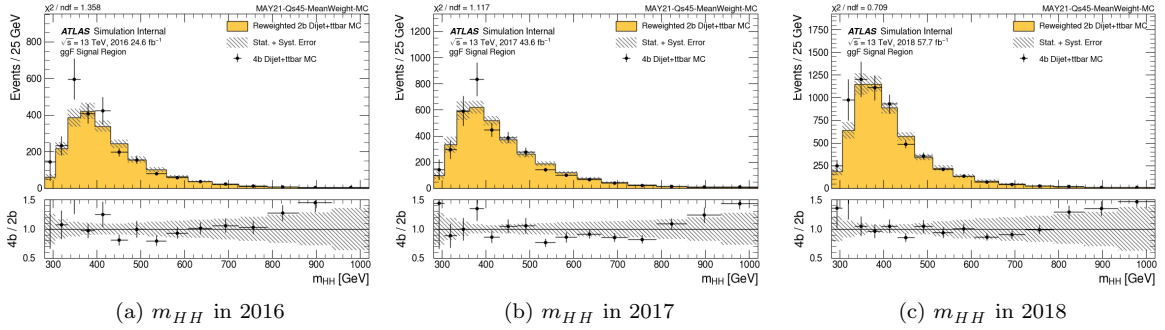


Figure 11.23: m_{HH} MC reweighted 2b events and 4b events for the QCD and $t\bar{t}$ MC samples. The background error includes the deep ensembles, 2b stat, and CR1/CR2 systematic errors.

This suite of background validation tests together gave us confidence moving forwards that we had a robust background estimate for trustworthy results.

12

Results

If we knew what we were doing, it wouldn't be research.

– Zhi's sweatshirt (Michigan physics department swag).

This chapter presents the $HH \rightarrow 4b$ results. In Chapter 12.1 the performance of the background model in predicting the $4b$ SR data is shown, while Chapter 12.2 briefly reviews the uncertainties affecting the signal model. Chapter 12.3 shows the results for the non-resonant signals in both the κ EFT frameworks, while Chapter 12.4 concludes with the impact of $4b$ in the ATLAS HH combination.

12.1 Background prediction

Figure 12.1 shows result the ggF background model fit to the $4b$ SR SR data (with μ fixed to 0), our “post-fit background model”. The ggF fit has separate categories for each year, but Figure 12.1 shows the summed contributions from all of the years. The data is in good agreement with the background model. The separate categories are shown in the individual plots, and also included are the SM and $\kappa_\lambda=6$ ggF signals. VBF background reweighting was derived inclusively over all the years and therefore the background fit and limits are also inclusive over the years. Figure 12.2 shows the post-fit for the VBF background model for the two $\Delta\eta_{HH}$ categories. Again good agreement is seen between the background and the data.

The corresponding data, post-fit background and signal yields in the defined analysis categories are shown in Table 12.1. For the SM ggF signal, the low X_{HH} low $\Delta\eta_{HH}$ are the most sensitive categories, while for the SM VBF signal, there is more signal in the high $\Delta\eta_{HH}$ category.

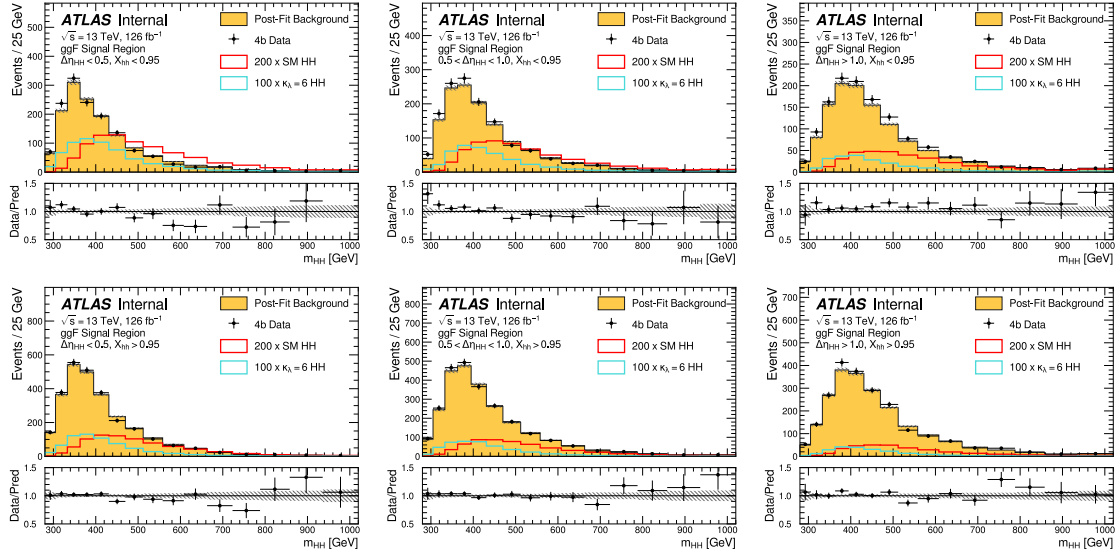


Figure 12.1: The post-fit background model for the ggF selection, with the predictions from all years combined. The post-fit error is shown in the dashed gray lines. The 4b data is shown in black, and the sub-panels show good agreement with the post-fit background prediction. Also shown are the ggF signals for the SM (scaled by a factor of 200) and $\kappa_\lambda=6$ (scaled by a factor of 100). Each plot is a separate kinematic category with the top row showing $X_{HH} < 0.95$, the bottom row showing $X_{HH} > 0.95$, and the columns showing the $\Delta\eta_{HH}$ categories.

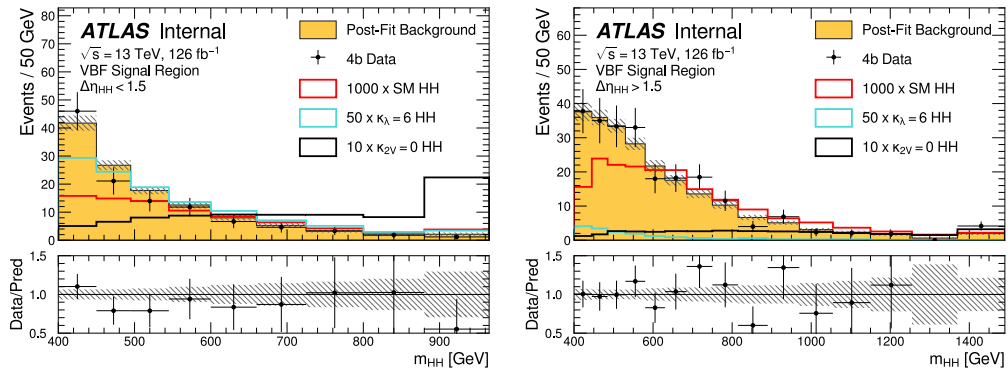


Figure 12.2: The post-fit background model for the VBF selection, with the post-fit error shown in the dashed gray lines. The 4b data is shown in black, and the sub-panels show good agreement with the post-fit background prediction. Also shown are the VBF signals for the SM (scaled by a factor of 1000), $\kappa_\lambda=6$ (scaled by a factor of 50), and $\kappa_{2V}=0$ (scaled by a factor of 10). Each plot is a separate kinematic category with $\Delta\eta_{HH} < 1.5$ on the left and $\Delta\eta_{HH} > 1.5$ on the right.

Category	Data	Bkg	ggF SM	VBF SM	$\frac{\text{ggF SM}}{\text{Bkg}}$	$\frac{\text{VBF SM}}{\text{Bkg}}$	
ggF signal region							
$X_{HH} < 0.95$	$\Delta\eta_{HH} < 0.5$	1940	1940 ± 130	7.0	0.038	3.6×10^{-3}	2.0×10^{-5}
	$\Delta\eta_{HH} \in (0.5, 1.0)$	1924	1870 ± 120	5.1	0.037	2.7×10^{-3}	2.0×10^{-5}
	$\Delta\eta_{HH} > 1.0$	1880	1740 ± 120	2.9	0.043	1.7×10^{-3}	2.5×10^{-5}
$X_{HH} > 0.95$	$\Delta\eta_{HH} < 0.5$	3602	3620 ± 200	6.5	0.036	1.8×10^{-3}	9.9×10^{-6}
	$\Delta\eta_{HH} \in (0.5, 1.0)$	3540	3490 ± 190	4.7	0.040	1.3×10^{-3}	1.1×10^{-5}
	$\Delta\eta_{HH} > 1.0$	3285	3210 ± 200	2.8	0.041	8.7×10^{-4}	1.3×10^{-5}
VBF signal region							
	$\Delta\eta_{HH} < 1.5$	116	125 ± 12	0.37	0.090	3.0×10^{-3}	7.2×10^{-4}
	$\Delta\eta_{HH} > 1.5$	241	231 ± 20	0.06	0.21	2.6×10^{-4}	9×10^{-4}

Table 12.1: The yields in each analysis category for the data, expected background (from the background only fit), and SM ggF and VBF signals [7].

12.2 Overview of signal systematic uncertainties

While the background model was entirely data-driven, the signals come from MC, and these samples have additional NPs which account for the known mismodellings of the simulation. Except for the custom E_T SFs (which were described in Chapter 10.1.3), these NPs follow the common procedure used across the collaboration, and are summarized below.

Detector modelling uncertainties

To account for the difference in the tagging probability in data versus MC, the dedicated FTAG SFs for the b , c , and light-jets are applied using the variations measured by the calibrations (described in Chapter 6.5). The variation on the jets' energy scale and resolution (JES and JER) are applied, as well as the uncertainty due to the pile-up veto JVT tagger. Since pile-up reweighting scale factors correct the simulation to the pile-up distribution in data, we apply the uncertainty for these reweighting factors. Trigger uncertainties from the HLT b -tag are prescribed by the b -jet trigger group and the dedicated E_T SFs derived for this analysis are also applied. The uncertainty on the integrated luminosity is accounted for as a 1.7% on the signal normalization uncertainty.

Theoretical Uncertainties

Differences due to assumptions for implementing the parton shower (PS) are assessed by comparing the Pythia 8 sample (which uses the Lund string model) with an alternative Herwig 7 sample (which uses the cluster hadronization model). This is the largest signal uncertainty with a 10% impact on the ggF and VBF acceptances. The $\pm 1\sigma$ variation templates are derived for each category, and treated as correlated in the fit.

The uncertainty in the matrix element is assessed by varying the renormalization and factorization

scales (μ_R and μ_F) up and down by a factor of 2. This uncertainty amounts to an $\approx 2\%$ difference for the signals, although it gets as large as 6% in some categories. The uncertainties due to the pdfs of the colliding partons are calculated using the 1σ error bar on a set of replicas sampled from the underlying CT14, MHT14 and NNPDF3.0 pdf sets [162].

For the limits on the SM signal strength, 3.5% normalization uncertainty is included for the $H \rightarrow b\bar{b}$ branching ratio, and the uncertainties on the cross-section due to the pdf, α_s , renormalization scheme, and m_t scale are accounted for with normalization uncertainty (with separate uncertainties for ggF and VBF production). The uncertainty due to the finite top mass in the ggF calculation is asymmetric (Chapter 9.1), so each production mode treats the top mass uncertainty with a separate normalization factor. The cross-section and branching ratio uncertainties are only included for the μ_{SM} limits, as the other results are given in terms of the experimentally measured cross-section.

The uncertainties and the assigned NPs are summarized in Table 12.2. The pulls for these signal and background NPs when fit to the observed data are shown in Figure 12.3. In the upper left plot showing the background pulls, the 2016 ggF background model has the largest pulls as this was the year where we saw the largest discrepancy between the pre-fit background model and the observed 4b data (see Figure C.6). But most of the background pulls are between ± 1 .

Systematic identifier	Description	# of NPs
Background	ggF: CR12 Extrapolation	3 years x 4 SR quadrants = 12
	VBF: CR1 / CR2 Shape variations	4 SR quadrants
	3b1f unc (ggF only)	3 years
	Bootstrap	# of histogram bins
FTAG	b -jet	3 eigenvariations and 1 extrapolation
	c -jet eigendecomp	4 eigenvariations and 1 extrapolation
	light-jet eigendecomp	4 eigenvariations
JET	JES	30
	JER	13
	JVT	1
Data taking	pile-up reweighting	1
	Luminosity	1
	HLT + L1 trigger SFs	2
Theory	PS, pdf, and scale uncertainties	3 (PS, pdf, scale) x 2 (ggF + VBF) = 6
	BR $H \rightarrow b\bar{b}$ (*)	1
	Cross-section uncertainties (*)	4

Table 12.2: Note: the (*) indicates these are only included in the μ_{SM} result.

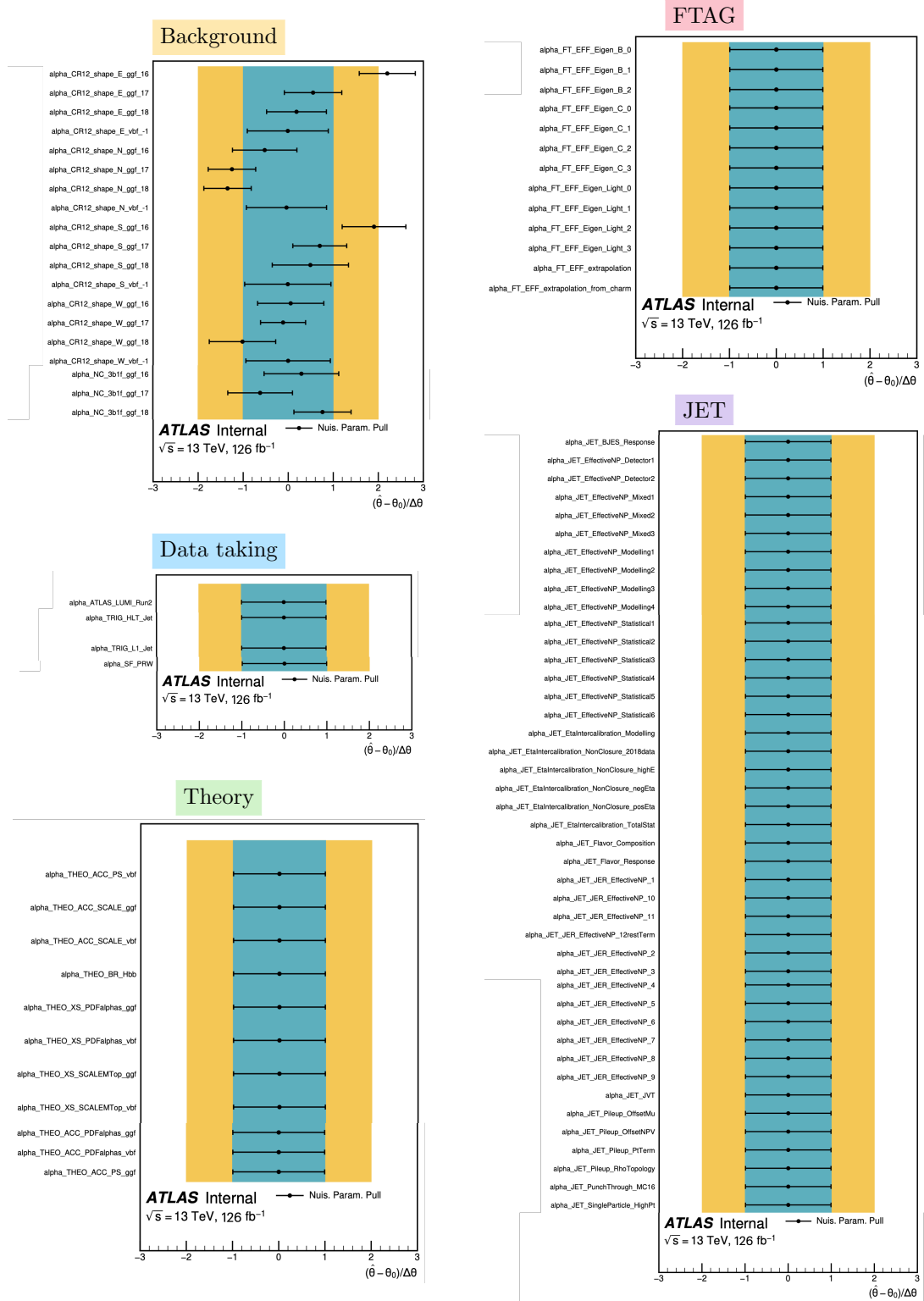


Figure 12.3: Pulls for the fit to the background template.

12.3 $HH \rightarrow 4b$ analysis results

The Parameter of Interest (PoI) for the scans is the signal strength $\mu = \sigma/\sigma_{SM}$, where the ggF and VBF channels are fit simultaneously with the same signal strength. The 95% confidence level limit on the SM signal strength is 5.4 (8.1) observed (exp). The limits just from the ggF analysis are 5.45 (8.09) observed (expected). The limit with only the statistical uncertainties included (the 2b stat uncertainty and signal mc stat uncertainty), is 6.0. If only the background systematic uncertainties are included, the limit is 7.1. Table 12.3 shows the impact of the dominant systematics on our SM limit. From the experimental side, the data-driven background systematics drive the sensitivity, with the deep ensembles + bootstrap uncertainty and the CR12 extrapolation systematic both giving an $\approx 7\%$ degradation in the final result. The current uncertainty on the theoretical HH cross-section has a 9% impact on our result. The observed (expected) upper limit on the ggF+VBF cross section is $5.02 \times 32.776 \text{ fb} = 164.49 \text{ fb}$ ($7.36 \times 32.776 \text{ fb} = 241.20 \text{ fb}$) when these theoretical cross sections are excluded.

Source of uncertainty	$\Delta\mu/\mu$
Theoretical	
Uncertainty on signal cross-section	-9.0%
All other theory uncertainties	-1.4%
Background modeling	
Deep ensembles + bootstrap	-7.1%
CR1 / CR2 extrapolation	-7.5%
3b1f non-closure uncertainty	-2.0%

Table 12.3: Impact of the various uncertainties on the $\mu_{ggF+VBF}$ limit. To test the impact of each systematic, it's profiled and then constrained to its best fit value as the upper limit is recalculated. Only the systematics that have an impact larger than a % are shown. All of the experimental uncertainties together had a sub-% level impact on the final result.

Figure 12.4(a) shows the result of the 95% confidence level upper limit on the $\sigma_{ggF+VBF}$ cross-section as κ_λ varies. At $\kappa_\lambda=6$ the observed limit has a positive fluctuation corresponding to $+1\sigma$. $\kappa_\lambda=6$ is one of the softest HH signals, and in the pre-fit background model plots in Figure C.6 shows that the background under predicts the yields at low m_{HH} . At high m_{HH} , the background model tends to overpredict the yield, which is why our limits are stronger than expected for the SM $\kappa_\lambda=1$ and the negative κ_λ values.

A constraint on κ_λ is extracted from the intersection of the cross-section limit with the theory prediction. Table 12.4 shows the κ_λ constraint, and compares to the predictions for the ggF and VBF channels separately. For the upper constraint on κ_λ , it is interesting that we get approximately equal strength from the ggF and VBF channels. For the lower κ_λ constraint, the ggF channel dominates, due to its much larger cross-section.

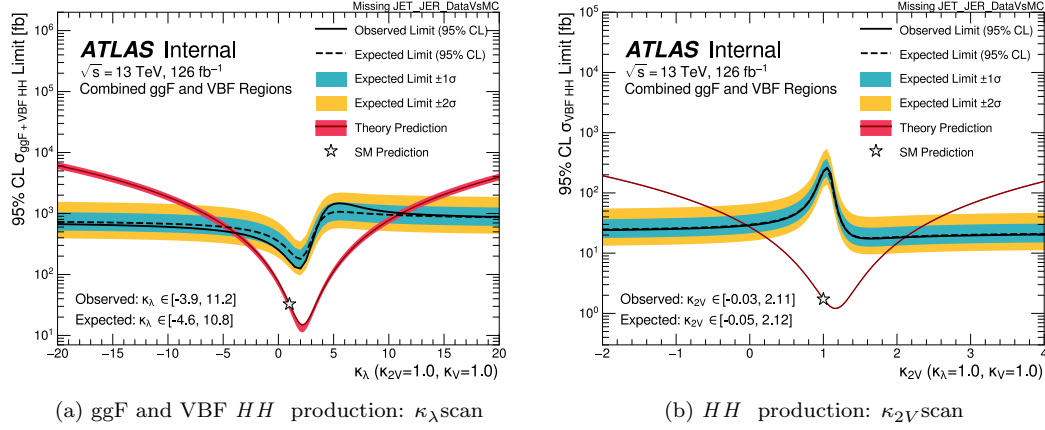


Figure 12.4: The 95% Confidence Level upper limits on the HH production cross-sections with respect to the κ_λ variations (left) and the κ_{2V} variations (right).

Channel	Observed Interval	Expected Interval
ggF Channel (ggF signal only)	$[-4.5, 13.3]$	$[-5.0, 12.0]$
VBF Channel (VBF signal only)	$[-10.0, 13.2]$	$[-12.3, 15.6]$
Combination (ggF+VBF signals)	$[-3.9, 11.2]$	$[-4.6, 10.8]$

Table 12.4: In the combined channel, the observed and expected limit intervals on the coupling modifier κ_λ at the 95% CL for the ggF channel, the VBF channel and the combination of the two.

The 95% confidence level upper limit on the σ_{VBF} cross-section for variations in the κ_{2V} coupling. For this result the SM ggF signal is treated as a background. There is good agreement between the observed and expected limits. The observed (expected) constraints on κ_{2V} are $[-0.03, 2.11]$ ($[-0.05, 2.112]$) – and we are very close to excluding the $\kappa_{2V}=0$ model.

Limits on the 2 dimensional $(\kappa_{2V}, \kappa_\lambda)$ and (κ_{2V}, κ_V) variations are shown in Figure 12.5.

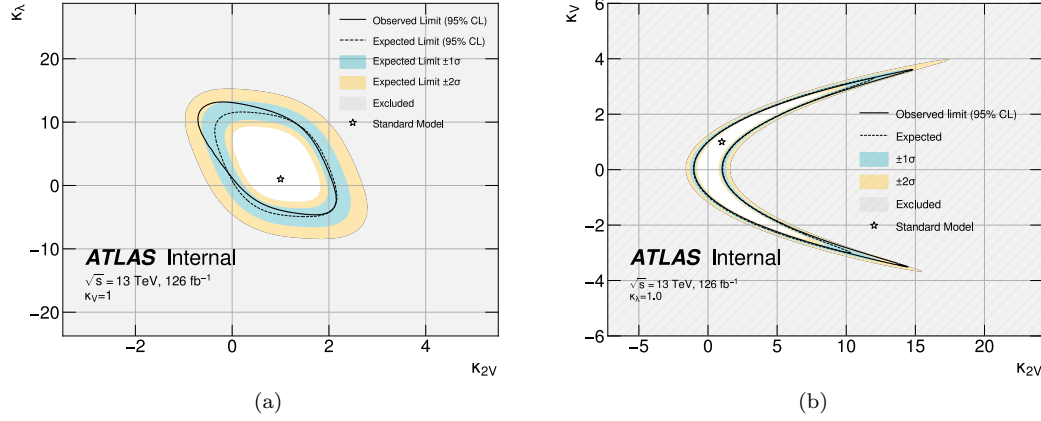


Figure 12.5: The observed (solid) and expected (dashed) limit intervals on the coupling modifiers κ_λ vs κ_{2V} (left) and κ_V vs κ_{2V} (right) at the 95% CL for the combination of the VBF+ggF channels. The solid lines show the couplings excluded by the observed 95% CL limits, while the dashed lines indicate the expected 95% CL limits. The blue (yellow) bands give the errors on the expected limits. The shaded grey regions are excluded. The stars indicate the SM couplings ($\kappa = 1$).

For the SMEFT and HEFT interpretations, only the ggF coupling variations are considered, and the limits are extracted just using the ggF channel. The 1-dimensional scans are performed for these five Wilson coefficients, and the constraints are reported in Table 12.5, where in each scan, all other coefficients are set to their SM values. The limits extracted for the seven BSM benchmarks in the HEFT parameter space are shown in Figure 12.6.

Parameter	Expected Limit	Observed Limit
$c_{H\Box}$	[-9.3, 14.8]	[-9.4, 13.9]
c_H	[-22.3, 10.5]	[-19.9, 10.8]
c_{tH}	[-11.0, 6.4]	[-10.1, 6.5]
c_{HG}	[-0.06, 0.06]	[-0.05, 0.05]
c_{tG}	[-1.13, 1.17]	[-0.96, 0.92]

Table 12.5: The extracted upper and lower limits on the SMEFT parameters to which the analysis is sensitive. For each parameter, the constraints are provided assuming the other parameters are fixed to 0.

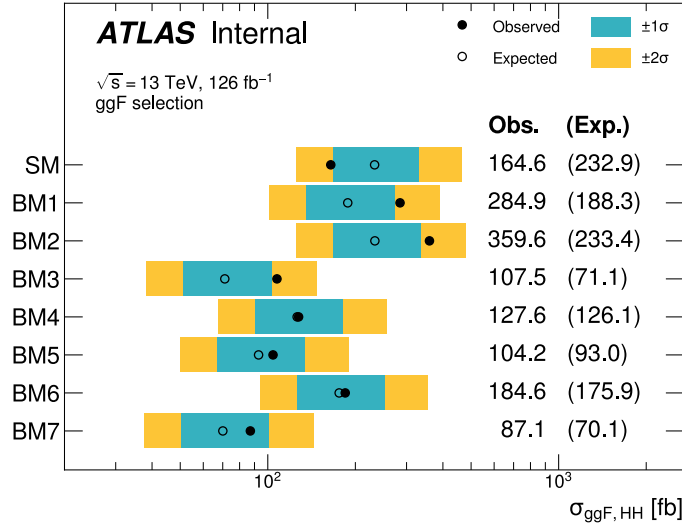


Figure 12.6: The extracted upper limits for the SM cross section, and the seven HEFT benchmarks.

12.4 Combination and analysis improvements

Figure C.7 shows the impact of combining these $4b$ results with the other main HH channels. Just combining the $b\bar{b}\tau^+\tau^-$ and $b\bar{b}\gamma\gamma$ gave an (already impressive) observed and expected limit on the SM signal strength of 3.1. Including our efforts from $4b$ improved on this combination by giving an observed (expected) limit of **2.9 (2.4)** – the *world limit* on the observed SM signal strength. The CMS combination combined more channels also including $b\bar{b}ZZ$, multileptons, and boosted $4b$, and has an observed (expected) strength on the SM HH signal of 3.4 (2.5) [136].

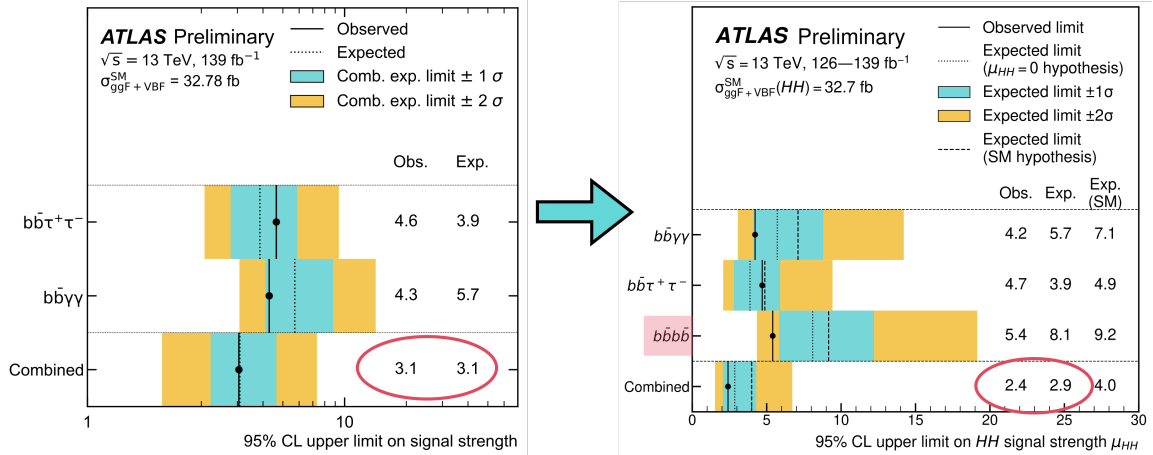


Figure 12.7: The SM limit combining the ATLAS channels [153] [163]

To give context for this latest $4b$ result, Figure 12.8 shows the results from our previous publications. Just with the extra statistics with increasing dataset size, we would have expected our upper limit to improve by $1/\sqrt{\mathcal{L}}$. (indicated by the grey line). However, these publication results are fitted to a curve $\mu = \kappa/\mathcal{L}^\alpha$, we get superluminous scaling with a coefficient of $\alpha = -0.76$, showing that we consistently exceed our expectations. The improvements in the latest result include the new DL1r tagger b -jet identification, and an improved analysis strategy with better background rejection and categorization to improve the power of the final discriminant.

Our hope is that through more creativity and innovation we will continue to improve this result beyond our expectations. The next chapter concludes by presenting one of these avenues for improvement relevant for the Run 3 iteration of this analysis.

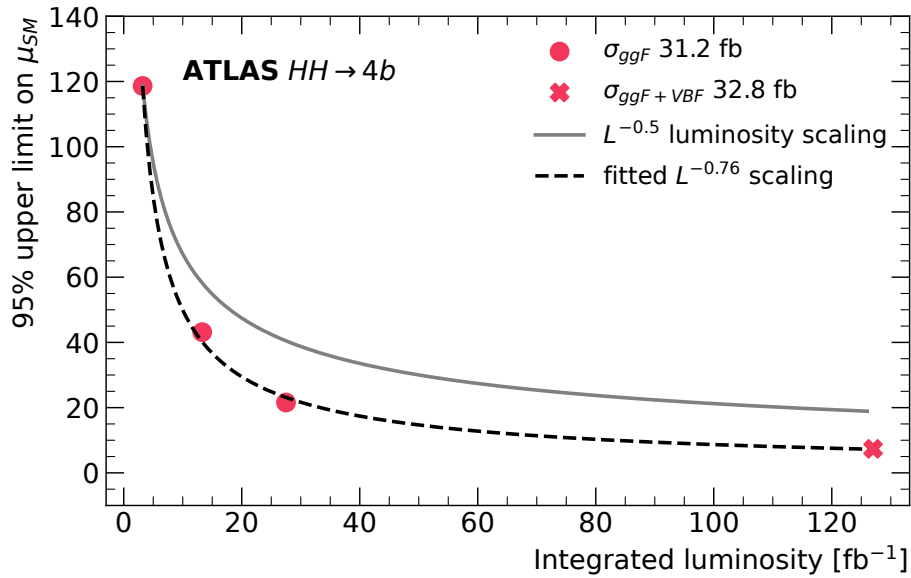


Figure 12.8: Run 2 publications of the expected $4b$ non-resonant limit, in units of the $HH \rightarrow 4b$ cross section [164], [165], [116], [163]. For a fair comparison across the publications, the theoretical cross-section uncertainties in [163], as these were not included in the earlier publications.

Part IV

Future developments

This final chapter presents one of the most promising future directions for the $HH \rightarrow 4b$ analysis from the ideas explored in this thesis work. A novel data-driven background estimation strategy is presented, and initial studies show an improved performance compared to the Run 2 background estimation described in Chapter 11.

13

Generative models for data-driven background modeling

Go with the flow. Force nothing. Let it happen, or not happen... trusting that whichever way it goes, it's for the best.

– Mandy Hale

As described in Chapter 11, QCD is 90% of the $4b$ background and needs to be estimated with data-driven techniques. In the Run 2 $4b$ analysis, reweighting maps were derived in dedicated CRs and then extrapolated into the SR. The uncertainty on this background estimate limits analysis sensitivity – in particular, the uncertainty from CRs extrapolating into the SR degrades the upper limit on μ_{SM} by 7.5% (Table 12.3).

This chapter proposes an alternative background estimate which relies on a different set of assumptions. Figure 13.1 shows that the multi-jet background processes vary smoothly across the (m_{H1}, m_{H2}) massplane. This smooth variation of the underlying physics is used as a key inductive bias to train a generative model as a function of (m_{H1}, m_{H2}) to pose the background estimate as a high dimensional interpolation. This is done with a two-step procedure where the Higgs candidate masses and the HH kinematics (given by the vector x) are modeled separately. The joint probability distribution is decomposed using the chain rule of probability as:

$$p(x, m_{H1}, m_{H2}) = p(x|m_{H1}, m_{H2}) \cdot p(m_{H1}, m_{H2}), \quad (13.1)$$

where x is a six-dimensional vector that models the rest of the Higgs candidates' kinematics (as will be elaborated on in Chapter 13.2.2). The $p(m_{h1}, m_{h2})$ is fit with a Gaussian process (GP, described in Chapter 13.1) to constrain the predictions inside of the SR using the correlation with events outside of the SR. Since GP does not scale well to high dimensional inputs [166], the rest of the event kinematics

x are modeled with a conditional normalizing flow: $p(x|m_{h1}, m_{h2})$ (described in Chapter 13.2). Since the smoothly varying (m_{H1}, m_{H2}) is a key modeling assumption, the interpolation is derived before applying the $X_{Wt} > 1.5$ cut which induced structures in the massplane.

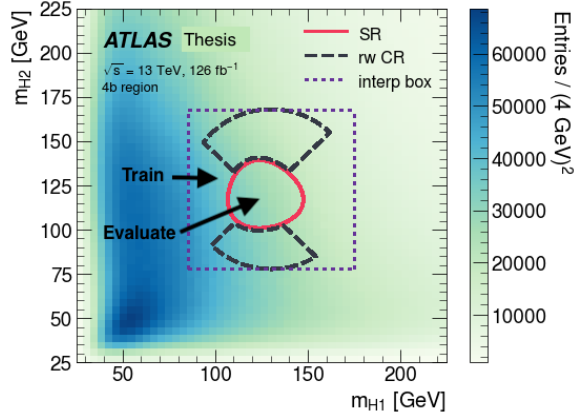


Figure 13.1: Illustration of the interpolation setup. The pink circle shows the nominal 4b SR, with the purple dotted line showing the bounding box used to train the interpolation. The quadrants for deriving the baseline reweighting method are shown in the navy blue dashed line crescents.

Here the ggF analysis background estimate is studied as this channel is more impacted by the background systematics than the VBF analysis which is dominated by statistical uncertainties. The selection here is identical to that outlined in Chapter 10, and the results for this interpolation are compared to the reweighting strategy in both the 4b SR, and the suite of validation regions.

13.1 Gaussian Processes

13.1.1 Introduction

The $p(m_{H1}, m_{H2})$ model fits a GP to the blinded 2d histogram, inspired by the boosted 4b analysis [117]. A GP is a Bayesian non-parametric ML method, which means that the model size grows with the number of input data points [167, 168]. It generalizes the notion of the multivariate Gaussian which describes distributions over random vectors, to describe the distributions over functions (to allow evaluations on arbitrary inputs) – where here we want to evaluate the distribution in the blinded SR. The GP specifies that the prior over any finite collection of points follows a multivariate Gaussian probability distribution, as specified by:

$$\begin{bmatrix} f(x_1) \\ \vdots \\ f(x_m) \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} m(x_1) \\ \vdots \\ m(x_m) \end{bmatrix}, \begin{bmatrix} K(x_1, x_1) & \cdots & K(x_1, x_m) \\ \vdots & \ddots & \vdots \\ K(x_m, x_1) & \cdots & K(x_m, x_m) \end{bmatrix} \right) \quad (13.2)$$

where x_1, \dots, x_m are the input points, and the f describes the prior over functions. As a pre-processing step, the targets are normalized to zero mean and unit variance, so the mean of the GP is set to zero. Therefore, the covariance function K defines the GP model. We will be using the squared exponential kernel for the covariance function: $K(x, x') = \exp\left(-\frac{(x-x')^2}{2l^2}\right)$, where l the length scale determines how strongly correlated close inputs are. The training data points be specified by the matrix X and the vector \mathbf{y} . Uncertainties on the i training observations can be included as:

$$y^{(i)} = f(x^{(i)}) + \varepsilon^{(i)}, \quad i = 1, \dots, m, \quad (13.3)$$

where ε is the noise $\varepsilon \sim \mathcal{N}(0, \sigma^2)$, and the (i) superscript denotes one of the i training examples. Since the GP specifies a prior over functions, this length scale defining K is optimized by maximizing the log marginal likelihood on the training inputs. The log marginal likelihood is:

$$\begin{aligned} \log p(\mathbf{y}|X) &= \log \frac{1}{(2\pi)^{n/2} |K + \sigma^2 I|^{1/2}} \exp\left(-\frac{1}{2} \mathbf{y}^T (K + \sigma^2 I) \mathbf{y}\right) \\ &= -\frac{1}{2} \mathbf{y}^T (K + \sigma^2 I) \mathbf{y} - \frac{1}{2} \log |K + \sigma^2 I| - \frac{n}{2} \log 2\pi, \end{aligned} \quad (13.4)$$

The term “marginal” means that the Gaussian distribution over the finite number of training data points doesn’t change if a larger set of data points is considered [167]. Maximizing $\log p(y|X)$ gives the maximum likelihood estimator for the length scale, l .

The joint marginal over the training points and the test (or interpolation) points X_* and \mathbf{y}_* also follows a multivariate Gaussian distribution:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{y}_* \end{bmatrix} \Big|_{X, X_*} = \begin{bmatrix} f \\ f_* \end{bmatrix} + \begin{bmatrix} \vec{\varepsilon} \\ \vec{\varepsilon}_* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K(X, X) + \sigma^2 I & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) + \sigma^2 I \end{bmatrix}\right) \quad (13.5)$$

Marginalizing out the dependence over the training variables gives the “predictive posterior distribution” as the (X_*, \mathbf{y}_*) also follows a Gaussian distribution, with the mean and covariance given by [167]:

$$\mu_* = K(X_*, X)(K(X, X) + \sigma^2 I)^{-1} \mathbf{y} \quad (13.6)$$

$$\Sigma_* = K(X_*, X_*) - K(X_*, X)(K(X, X) + \sigma^2 I)^{-1} K(X, X_*). \quad (13.7)$$

Instead of just having a point prediction at each query point, there is a distribution described by this multivariate Gaussian – which allows the model to also have a measure of its uncertainty.

13.1.2 Application for the $HH \rightarrow 4b$ analysis

The input to the GP is a 2d histogram of the massplane with a bounding box around the circle that defined the reweighting CRs, as shown by the dotted purple line in Figure 13.2(a). This box has an edge length of 90 GeV, and 25 bins for both the m_{H1} and m_{H2} axes. The (m_{H1}, m_{H2}) bin centers are used to predict the number of events in the bin. The Poisson errors on the bin entries are the uncertainties in the regression targets, and are included in the GP fit with the σ in Eq. 13.3. The targets are normalized to have zero mean and unit variance. Bins with any overlap with the SR are not included in the training process.

The GP is fit with scikit-learn [166] using the squared exponential kernel with two length scales for the m_{H1} and m_{H2} directions:

$$K \left(\begin{bmatrix} m_{H1} \\ m_{H2} \end{bmatrix}, \begin{bmatrix} m'_{H1} \\ m'_{H2} \end{bmatrix} \right) = \exp \left(- \frac{(m_{H1} - m'_{H1})^2}{2l_1^2} - \frac{(m_{H2} - m'_{H2})^2}{2l_2^2} \right). \quad (13.8)$$

To account for the differences in the triggers, a separate GP is fit for each year (2016, 2017, and 2018). Figure 13.2(b) shows the mean of predictive posterior for the GP fit to the blinded massplane. The predicted massplane gives a smoothed prediction of the input massplane, while also providing an interpolation for the bin entries into the SR since the fitted length scales (shown in Table 13.1) are larger than the radius of the SR. The GP predictions are constrained by nearby observed data points, so Figure 13.2(c) shows that the GP predicted error increases slightly as it interpolates into the SR. To quantify the error compared with the observed data, Figure 13.3 shows pulls defined by the GP mean prediction minus the observed yield, divided by the Poisson uncertainty on the observation (the square root of the bin yields). The pulls for these bins lying outside (purple) and inside (pink) the SR is shown in Figure 13.3(b). The mean and variance of these pulls in the SR are: $\mu = 0.08$, $\sigma = 0.93$ – which is close to zero mean and unit variance. This indicates the model is providing a good interpolation into the SR.

	m_{H1} [GeV]	m_{H2} [GeV]
2016	108.3	50.6
2017	106.1	64.4
2018	105.4	56.4

Table 13.1: The fitted length scales for 4b data.

Samples from the GP are used to condition the normalizing flow model’s SR prediction. Inverse transform sampling [169] is used to draw the samples of the 2d histogram bin centers. The two-dimensional (m_{H1}, m_{H2}) GP prediction is reshaped to a 1d histogram which is used to construct the cumulative density function (CDF). Samples of $z \sim [0, 1]$ are mapped to the bins of the CDF to give samples of the (m_{H1}, m_{H2}) bin centers following the GP’s $p(m_{H1}, m_{H2})$ probability. The (m_{H1}, m_{H2}) samples are then smeared uniformly with the bin width to give continuous predictions.

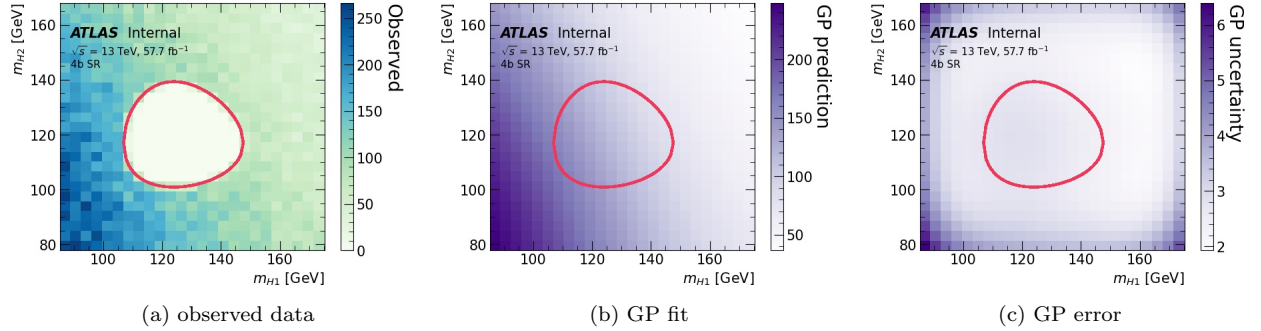


Figure 13.2: GP fits for the 2018 4b SR. All the massplane fits are before the X_{Wt} cut.

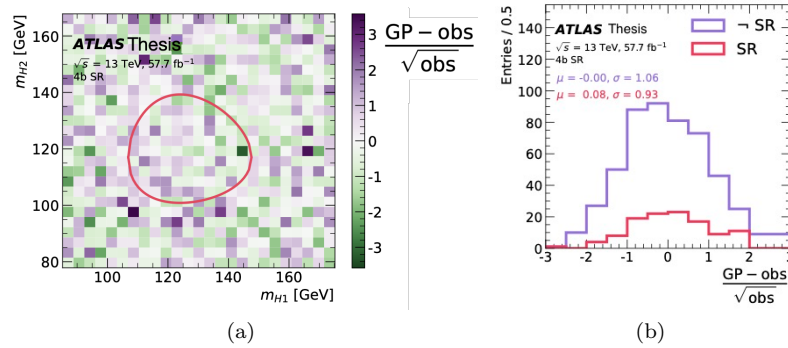


Figure 13.3: Left: non-closure of the GP fits for the 2018 4b SR region. Right: 1d histogram of the pulls. The “¬ SR” (“not SR”) purple line shows the bins used to fit the GP, while the “SR” pink line shows the bins that were blinded in the fit. These evaluation plots are before the X_{Wt} cut.

Since the number of samples drawn is arbitrary, the predicted SR event yields are rescaled using the observed and sampled entries from the interpolation bounding box with the blinded SR:

$$n_{SR} = n_{\neg SR}^{obs} \cdot \frac{n_{SR}^{samples}}{n_{\neg SR}^{samples}}. \quad (13.9)$$

13.2 Normalizing Flows

13.2.1 Intro

A normalizing flow is a generative model that transforms a known base density (such as a Gaussian) into a target density by a sequence of invertible transformations, here denoted f_i [170]. In the forward mode of the flow ($f = f_1 \circ \dots \circ f_L$ in Figure 13.4) samples from the Gaussian are transformed to give samples from the target distribution. Since each of the steps of the flow steps are invertible, the reverse mode of the flow f^{-1} lets us evaluate the probability of the samples using the chain rule of

probability:

$$p_x(x) = p_z(z) \left| \frac{dz}{dx} \right| = p_z(f^{-1}(x)) \left| \frac{df^{-1}}{dx} \right| \quad (13.10)$$

where $p_z(z)$ is the Gaussian base density, and x are the features we want to model. By estimating the probability of training data points, we can optimize the flow’s parameters with maximum likelihood. We define a loss as the negative log-likelihood of the training data ($-\sum_i \log p_x(x_i)$) and minimize it by stochastic gradient descent.

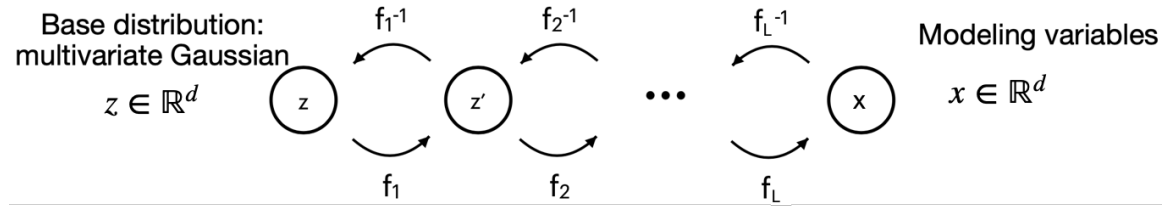


Figure 13.4: Visualization of the normalizing flow paradigm. The **forward mode** ($f = f_1 \circ \dots \circ f_L$) generates samples. The **reverse mode** ($f^{-1} = f_L^{-1} \circ \dots \circ f_1^{-1}$) evaluates the density of a data point.

Since $x \in \mathbb{R}^d$, the f_i transformations need to keep both the Jacobian and the f_i^{-1} calculations tractable. Various methods for flow architectures to maximize the expressivity while preserving tractability are described in recent reviews [171, 172]. A trick is to use “coupling layers” which only transform half of the variables at a time. The Jacobian becomes block diagonal and the inverse operation goes from $\mathcal{O}(d^3)$ to $\mathcal{O}(d)$ since we just need to take the product over the diagonal entries. The RealNVP is a simple architecture with such coupling layers used in our early optimization studies, and described in Appendix G.1.

Here we show results for a rational quadratic neural spline flow (RQ-NSF) [173] with their `nflows` python package [174]. Since the only constraint needed to ensure the invertibility of f_i is that the f_i are monotonic functions, this method uses a monotonically increasing function parametrized by a spline with K bins between (B, B) as shown in Figure 13.5. Each “bin” on the domain is parametrized by a rational quadratic transformation given by:

$$f_{jk}(x_i) = \frac{a_{ijk}x_i^2 + b_{ijk}x_i + c_{ijk}}{d_{ijk}x_i^2 + e_{ijk}x_i + f_{ijk}}$$

where i is the dimension of the transforming variable, and j be the corresponding flow layer, and k is the bin. To predict one step of a single transforming variable involves $6K$ constants: $a_{ijk}, b_{ijk}, c_{ijk}, d_{ijk}, e_{ijk}, f_{ijk}$.

- Boundary conditions:
 $(x_0, y_0) = (-B, -B)$ and $(x_K, y_K) = (B, B)$
- Continuity at the internal knots (which specify the bin boundaries):
2(K-1) constraints
- Continuous derivatives at the knots:
K-1 constraints

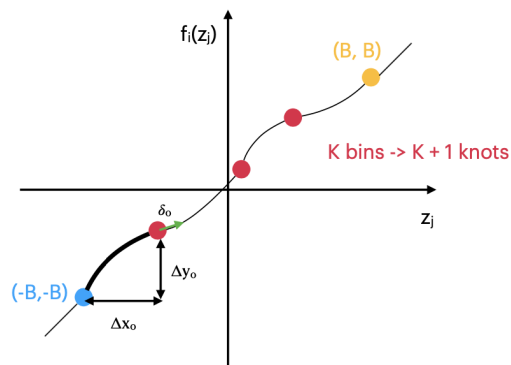


Figure 13.5: RQ-NSF parametrization

reduces the number of constants to

$$6K - [2(K-1) + 2K+1 + 2 + 2] = 3K - 1.$$

A NN then predicts the $2K$ widths and heights of the bins and the $K - 1$ derivatives at the internal knots. The NN output is an unconstrained real vector in $\theta_i \in \mathbb{R}^{3K-1}$, and this the θ_i vector is partitioned into three pieces $[\theta_i^w, \theta_i^h, \theta_i^d]$. Since the spline characterizes a transformation from a domain $(-B, B)$ into a range $(-B, B)$, the θ_i^w and θ_i^h vectors are each passed through a softmax function (with a range $(0,1)$) and then multiplied by $2B$ to give the widths and heights of the shifts between the knot locations. Since the monotonicity of the f_j is crucial for its invertibility, the θ_i^d representing the knot derivatives are passed through a softplus¹ function to ensure the derivative stays positive. Outside of the range $[-B, B]$ an identity transform is used. A batch norm layer [124] is used between each of the flow steps to keep the modeling variables in the range where the spline has its expressive power.

To allow mixing between the input variables, this architecture also uses a generalized permutation:

$$W = PLU, \quad (13.11)$$

where P is the permutation matrix, L is a lower triangular matrix, and U is an upper triangular matrix. The Jacobian stays tractable because $\det(W)$ only takes $\mathcal{O}(d)$ time to compute. Inverting this step involves solving two triangular systems, which is a $\mathcal{O}(d^2)$ operation, which is the same order as inverting the spline transformations.

To solve the interpolation problem, we predict the conditional probability distribution (as shown in [173]) by passing (m_{h1}, m_{h2}) as additional variables to the NNs predicting the spline constants. This preserves the nice properties of the bijection for each of the flow steps while keeping the Jacobians tractable. The flow is trained inclusively over the years (2016, 2017, and 2018), and the year is passed as an additional input.

¹Softplus(x) = $\log(1 + \exp(x))$: a smooth approximation to the function that ensures positive outputs [175].

13.2.2 Implementation details

We found that it was easier to use the flow to model the Higgs Candidates' 3-momentum instead of modeling m_{HH} and $\Delta\eta_{HH}$ directly. The top row of Figure 13.6 shows the HC kinematics. We predict the log of the HC p_{Ts} instead of the HC p_{Ts} directly since this gives a bell-shaped distribution which is easier to model for the flow which transforms a base Gaussian distribution. In Figure 13.6(c), the HC ϕ s are shown. If the HC ϕ were predicted directly, then the flow would not be able to tell that $\phi = \pi$ and $\phi = -\pi$ correspond to the same event. But the boundary condition is not the only aspect that we need to encode as the problem can be further simplified with the azimuthal symmetry in our events. We don't need to predict both ϕ_{H1} and ϕ_{H2} , but just the difference $\Delta\phi_{HH}$ as both m_{HH} and $\Delta\eta_{HH}$ are symmetric with respect to azimuthal rotations of the event. Figure 13.6(d) shows $\Delta\phi_{HH}$, which is peaked at π for the frequently back-to-back the di-jet pairs. The sharp cutoff would be very hard for the flow to predict with the smooth transformations from a base Gaussian distribution. We instead consider $\pi - \Delta\phi_{HH}$, to get a steeply falling distribution, and then take the log to model $\log(\pi - \Delta\phi_{HH})$. This transformation is more Gaussian-like, as shown in Figure 13.6(e). This transformation also encodes a physicality requirement that the predicted $\Delta\phi_{HH}$ are always less than π . There's no corresponding guarantee that the predicted $\Delta\phi_{HH} > 0$, but since this is in the tail of the distribution, the flow predicts a negligible number of negative $\Delta\phi_{HH}$ values (see Figure 13.10(f)). All modeling variables are normalized to have zero mean and unit variance.

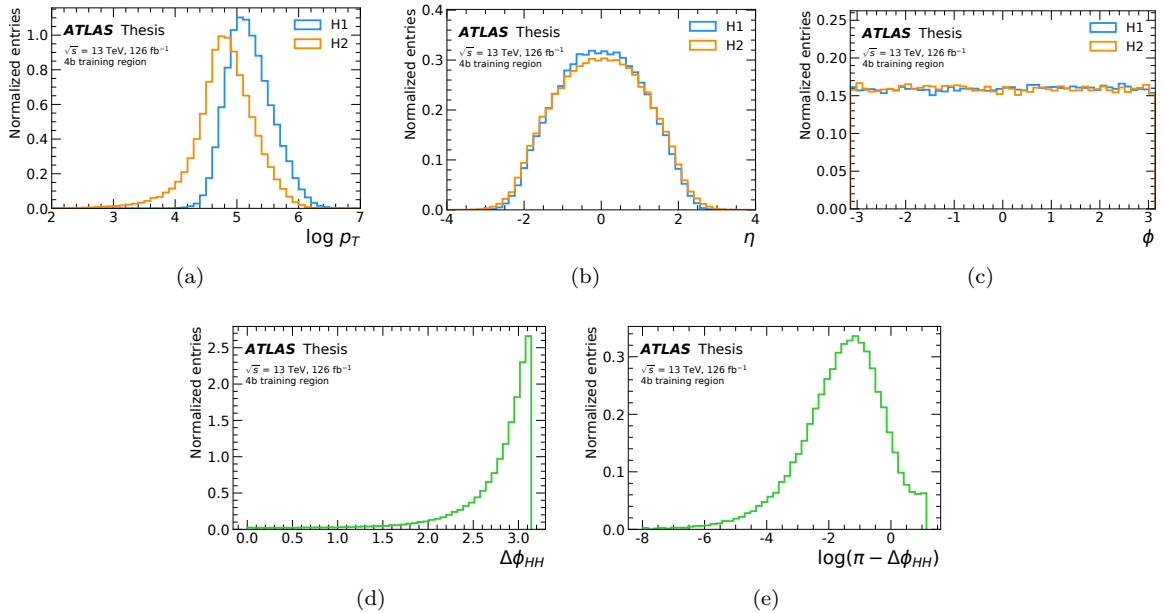


Figure 13.6: Motivation for Higgs Candidate variables modeled by the flow. The flow models the Higgs Candidates' p_{Ts} (a) and η s (b), the $\log(\pi - \Delta\phi_{HH})$ (e), and X_{Wt} (not shown here).

The a smooth massplane was a key modeling assumption, but the X_{Wt} cut induces structure in the massplane (see Figure 10.18). Thus, the flow is trained before applying the X_{Wt} cut, but X_{Wt} is passed as an additional modeling variable. When evaluating the SR predictions, the sampled events that have $X_{Wt} < 1.5$ are removed to compare to the reweighting background estimate.

We optimized the hyperparameters for this problem by looking at the modeling in the 2b SR, with Appendix G.2.3 giving the hyperparameter scan considered. We use a neural spline flow with 10 layers (or 10 f_i spline transformation steps). Each spline used $K = 4$ bins, and the spline transform is defined on the domain $(-B, B) = (-3, 3)$. The ResNets [176] are used to predict the spline constants. We use only needed a single ResNets block with 32 hidden units, and a dropout fraction of 10% [110]. To avoid overfitting, we found that using an L2 regularization with $\beta = 1e - 6$ was helpful, so the loss function is the sum of two terms:

$$\mathcal{L} = -\frac{1}{N} \sum_{x \sim \neg \text{SR}} \log p(x|m_{H1}, m_{H2}, \text{yr}) + \beta \sum_j w_j^2, \quad (13.12)$$

where the first term is the negative log-likelihood of the training data, and the second term is the L2 regularization for the w_j NN weights.

The flow was trained inclusively on all of the years using the data events in the interpolation bounding box outside of the SR conditioning on m_{H1}, m_{H2} and the year. We trained with the adam optimizer with a learning rate of 10^{-3} . Each flow is trained 25 times to assess an uncertainty from the variation in the NN trainings [127].

Figure 13.7 shows how flow proceeds in 10 steps to gradually transform the base Gaussian density into the Higgs candidate variables in the 4b SR.

Figure 13.8 visualizes the multi-dimensional density that the flow has learned by showing all of the $2d$ correlations. The grey scatter plots in the lower left triangular block show the correlations in the 4b SR data, while the pink scatter plots in the upper right triangular block show the analogous correlations predicted by the flow samples. The histograms along the diagonal compare the one-dimensional marginals. In the blue box the η_{H1}, η_{H2} correlation is emphasizes that the flow has learned to predict the $\Delta\eta_{HH} < 1.5$ cut.

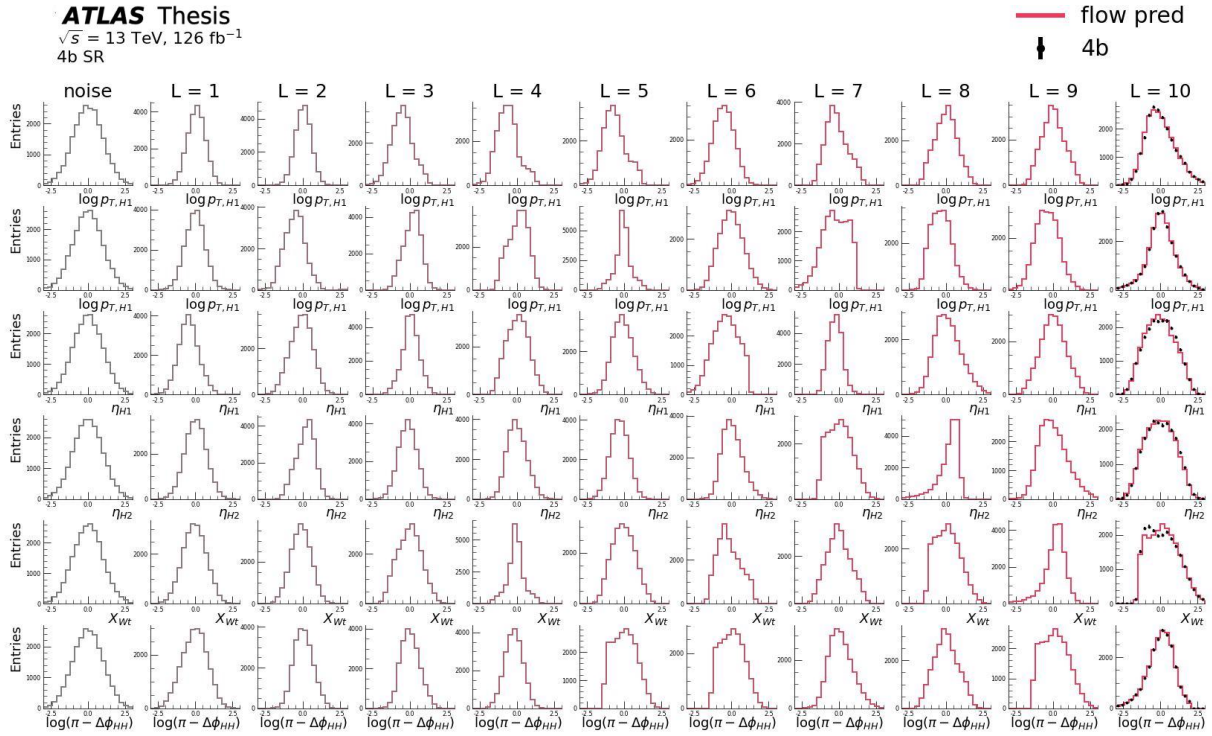


Figure 13.7: Demonstration of how the flow transforms unstructured noise into a structured prediction in the 4b SR. These samples are conditioned on the SR $(m_{H1}, m_{H2}, \text{yr})$ data from the 2016, 2017 and 2018 datasets. The grey histograms in the left column are samples from a 6d Gaussian. Each column to the right shows the transformation from one layer of the flow (i.e., a single invertible transformation), with the right-most column showing the prediction of the final flow. This is a single flow training, with the distributions are shown before applying the X_{Wt} . Variables have been scaled to zero mean and unit variance.

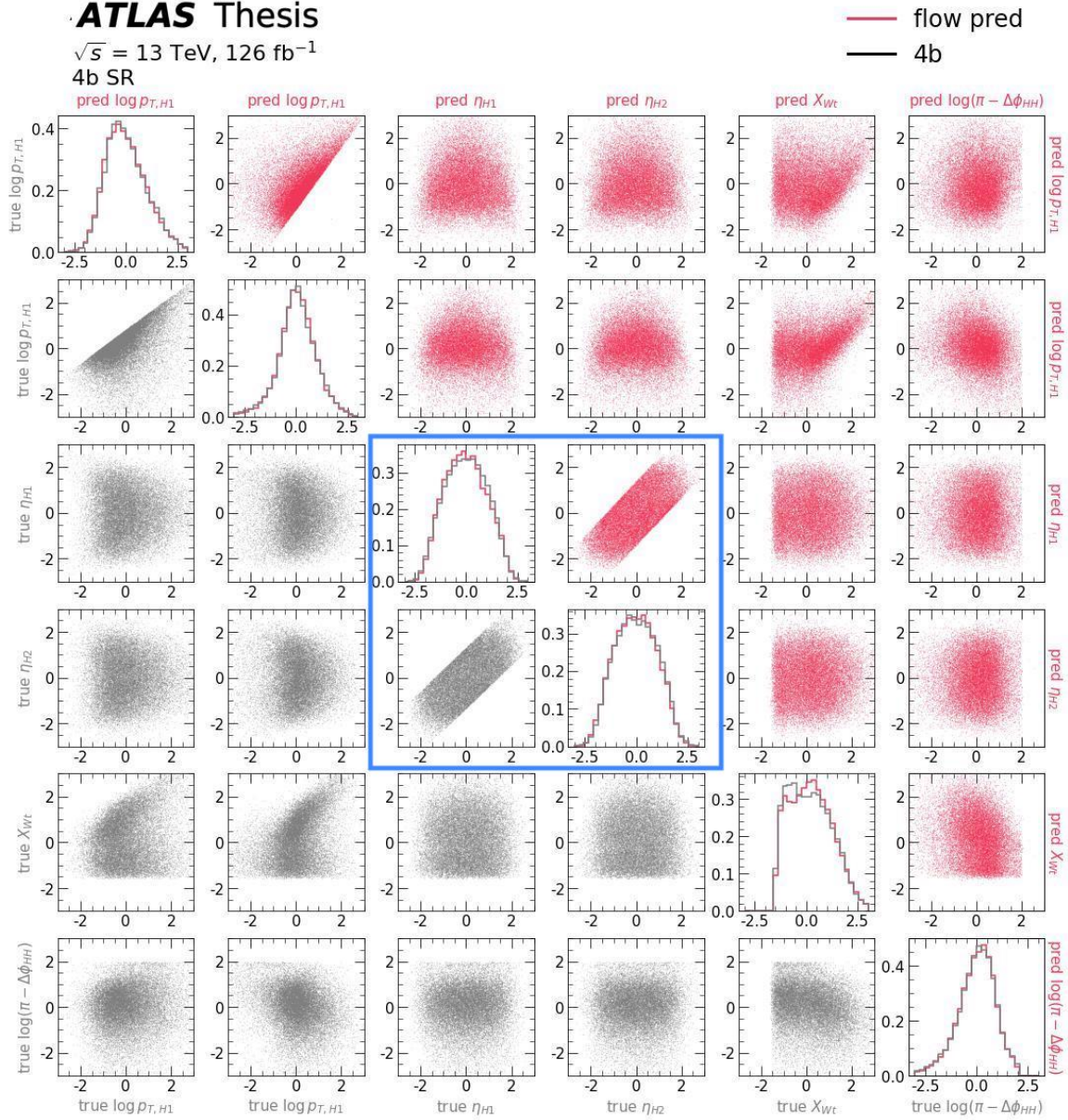


Figure 13.8: The correlation between the modeling variables for the SR data (grey) and the flow prediction (pink) in the 4b SR. The blue box emphasizes the η_{H1}, η_{H2} variables to show how the flow has learnt the $\Delta\eta_{HH} < 1.5$ cut. The marginal plots along the diagonal are normalized to unity. These samples are conditioned on the SR ($m_{H1}, m_{H2}, \text{yr}$) data from the 2016, 2017 and 2018 datasets. This is a single flow training, with the distributions shown before applying the X_{Wt} cut. Variables are scaled to zero mean and unit variance.

After predicting the vector $x = (\log p_{T,H1}, \log p_{T,H2}, \eta_{H1}, \eta_{H2} \log(\pi - \Delta\phi_{HH}), X_{Wt})$ from the flow, we combine this with the m_{H1}, m_{H2} samples from the GP to reconstruct the HC 4-vectors:

$$p_{H1} = (p_{T,H1}, \eta_{H1}, 0, m_{H1})$$

$$p_{H2} = (p_{T,H2}, \eta_{H2}, \Delta\phi_{HH}, m_{H2}).$$

These HC 4-vectors are then summed to get the m_{HH} , and we also calculate $\Delta\eta_{HH} = |\eta_{H1} - \eta_{H2}|$.

Figure 13.9 is a summary of the background model presented in the previous two sections. In evaluating the background model, 100,000 events are sampled from the interpolation box for each of the GP massplanes for each year: $p_{16}(m_{H1}, m_{H2})$, $p_{17}(m_{H1}, m_{H2})$, $p_{18}(m_{H1}, m_{H2})$, where the subscripts denote the 2016, 2017, and 2018 GP fits, respectively (step 1a in Figure 13.9). These samples set the SR normalization (step 1b and Eq. 13.9), and are used to condition the flow samples drawn in the SR (step 2). Although by modeling probability distributions both the GP and the flow quantify an uncertainty, for the 4b background modeling the dominant uncertainty was the Deep Ensembles uncertainty from the random initialization and optimization of the NNs ([127] and Table 12.3). For the following results, we show the NSF flow model trained 25 times. The same m_{H1}, m_{H2} samples from the GP SR are used to evaluate the flow prediction, and we cite the standard deviation of these 25 flow models as the interpolation error bar.

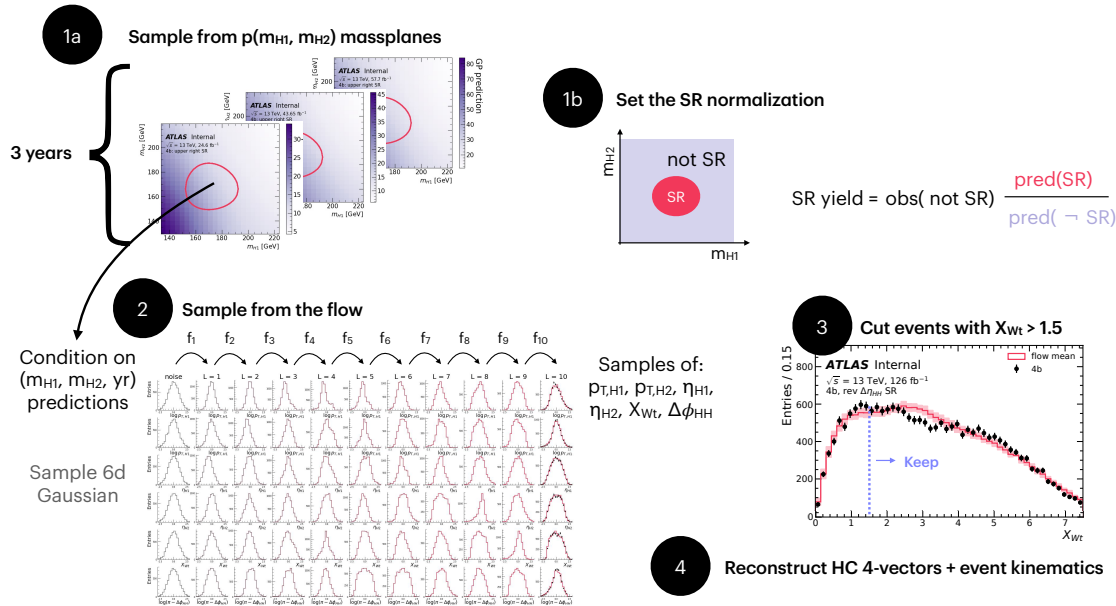


Figure 13.9: Demonstration of the background prediction algorithm for the upper right SR.

Comparisons are made to the reweighting method, and the reweighting error bars include the

deep ensembles error, the error on the CR1 / CR2 shape difference systematic, and the $2b$ statistical uncertainty.

13.3 Results

Figure 13.10 shows the six variables that the flow models and compares to the reweighting prediction. For the HC $p_{T,S}$, the flow is doing a better job modeling the peak of the distribution. Through careful flow input optimizations, we have also learned non-trivial distributions, such as $\Delta\phi_{HH}$ in Figure 13.10(f). The interpolation also is better at modeling the decreasing X_{Wt} distribution in the range $X_{Wt} = (3, 6)$.

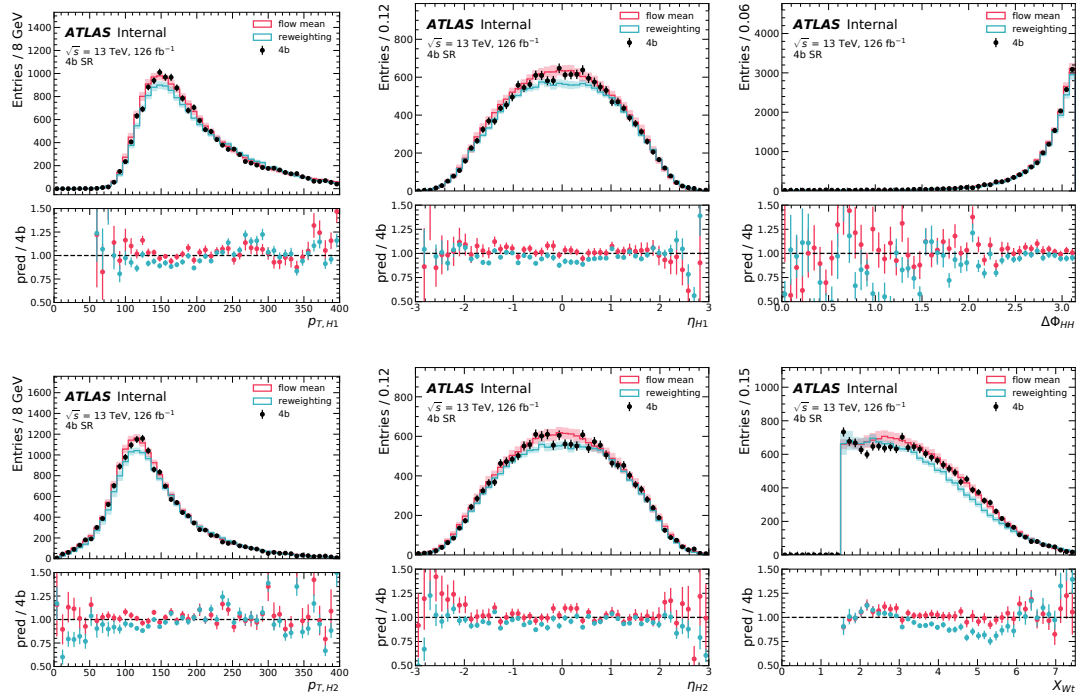


Figure 13.10: The flow training variables in the 4b SR (after the X_{Wt} cut).

Figure 13.11 shows the modeling for for the high level variables reconstructed from the HC kinematics. For both m_{HH} and $p_{T,HH}$ we see a better modeling in the peaks of these distributions. For $\Delta\eta_{HH}$, the flow sometimes predicts events with $\Delta\eta_{HH} > 1.5$, and in Figure 13.11(c) these overflow events are included in the highest $\Delta\eta_{HH}$ bin.

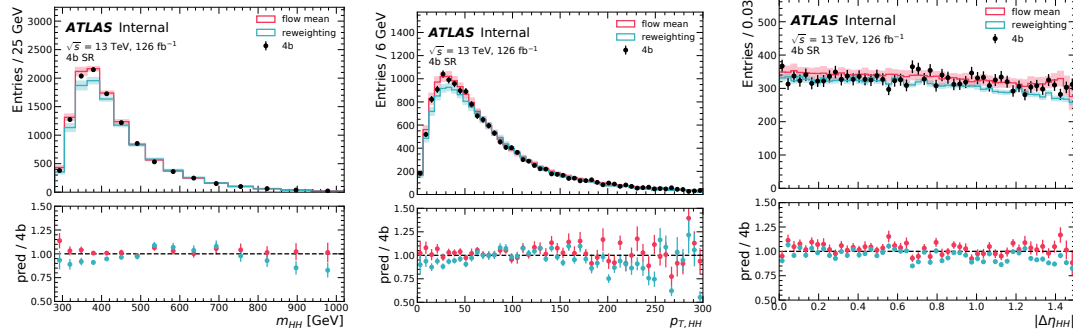


Figure 13.11: High level variables reconstructed from HC kinematics in the 4b SR (after the X_{Wt} cut).

Finally, Figure 13.12 shows m_{HH} with the $\Delta\eta_{HH}$ and X_{HH} categories used as the discriminating variable for setting the 4b analysis limits. Events in the overflow $\Delta\eta_{HH}$, m_{HH} bins are included in the highest bins here again. The flow again compares favorably here – making this a promising method to use for the Run 3 iteration of the 4b analysis.

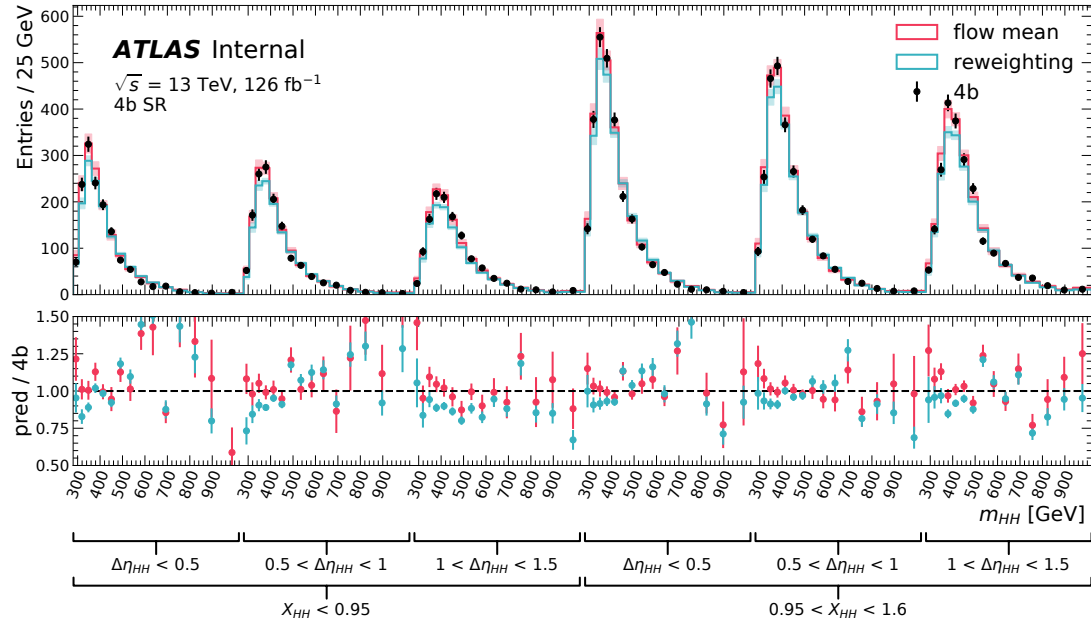


Figure 13.12: High dimensional discriminant in the 4b SR, after the X_{Wt} cut.

Results for other background validation regimes

Since one of the achievements of the recent 4b result was the tour de force efforts for testing our procedure in a suite of background validation regions, for the interpolation we also evaluated our procedure on this same set of regions: 3b1f, reversed $\Delta\eta_{HH}$, and the shifted SRs. Figure 13.13 shows the shifted SRs considered for the interpolation comparisons. Here we additionally also show results for the challenging “lower left” SR.

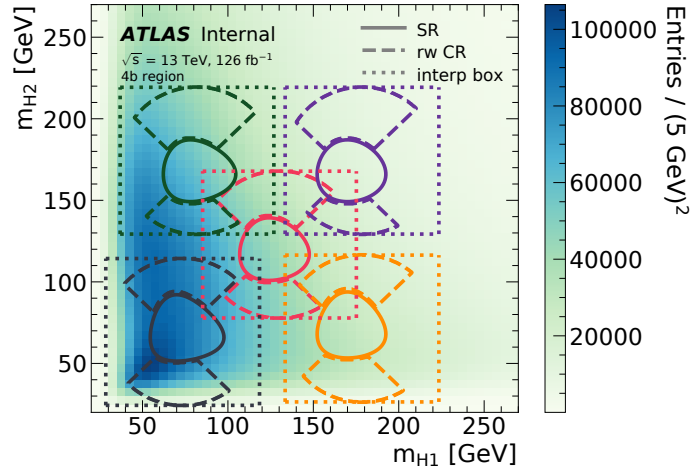


Figure 13.13: Illustration of the interpolation regimes. The pink (solid) circle shows the nominal 4b SR, with the pink dotted line showing the interpolation bounding box. The quadrants used to define the reweighing are also shown in the pink dashed crescents. The shifted regions used as validation tests of the method are shown in the blue, orange, green, and purple overlays.

Table 13.2 shows the observed yield, the reweighting, and the interpolation predictions. The flow enters in the interpolation’s yield prediction by modeling X_{Wt} as these results are shown after X_{Wt} cut. In pink are highlighted the regions the flow is more accurate at predicting the normalization, while the turquoise indicates the regions the reweighting does better.

	obs	rw	flow	1 - rw / obs [%]	1 - flow / obs [%]
lower left	40578	48708.9	39252.2	-20.0	3.3
lower right	12377	14648.5	11982.7	-18.4	3.2
upper right	5751	5543.0	5825.9	3.6	-1.3
upper left	19075	19504.7	19833.4	-2.3	-4.0
3b1f	180044	175817.9	175416.8	2.3	2.6
rev $\Delta\eta_{HH}$	16113	16462.7	16185.9	-2.2	-0.5
4b	16171	15423.7	16564.8	4.6	-2.4

Table 13.2: Yields for the predictions after applying the X_{Wt} cut

The m_{HH} modeling for each of these validation regions is also shown in Figure 13.14. Figure 13.14(a) shows the lower left SR. In the reweighting validation, we excluded this from the reweighting tests because we believed that it was more challenging than the 4b SR – as Figure 13.13 shows that the lower left CR1 (in the dashed navy crescents) overlaps with the kinematic turn-on in the massplane. It is more difficult to derive a reliable background estimate when the underlying control region does not smoothly vary into the SR, evidenced in Figure 13.14(a) as the non-closure is much larger than the assessed error. The interpolation model, however, is showing a %-level error for the non-closure in this challenging extrapolation regime.

The lower right SR in Figure 13.14(b) also has a larger non-closure for the reweighting compared to the interpolation. In this region, we understood that the reweighting did have a good closure for CR2. Since the error bar on the reweighting includes the variation from the alternative CR2 model, we can see that the CR2 provides a good model as the lower edge of the reweighting error bar is close to the observed data. There is still a benefit with the interpolation not needing to assess as large of an error bar.

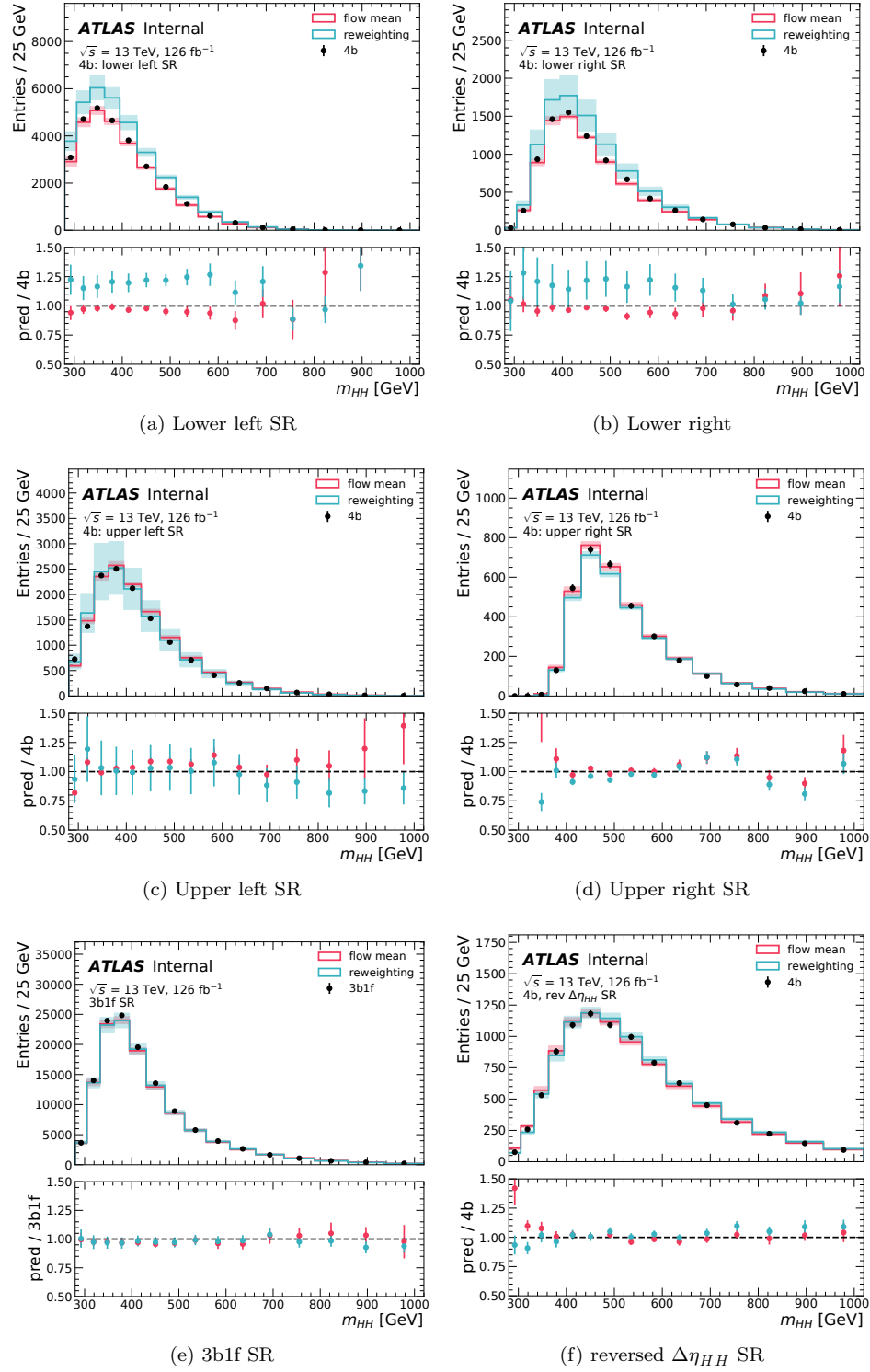
For the upper left SR in Figure 13.14(d), the reweighting has a better normalization prediction than the interpolation, but since the difference between CR1 and CR2 is quite large for this region, this is reflected in a correspondingly larger error bar for the reweighting.

The reweighting also has a slightly better normalization in the 3b1f region, but Figure 13.14(e) shows that both of these models give a good background model.

Finally, in Table 13.3 the χ^2 for the shapes of the kinematics are compared for the other validation regions. The prediction histograms are normalized to the observed event yield. The same highlighting scheme is used here as earlier, and overall the flow is showing good modeling in these regions with respect to the strong baseline provided by the reweighting.

	lower left		lower right		upper right		upper left		3b1f		rev $\Delta\eta_{HH}$		4b	
	rw	flow	rw	flow	rw	flow	rw	flow	rw	flow	rw	flow	rw	flow
3d disc	5.52	2.19	1.32	0.98	1.35	0.93	2.54	1.60	1.85	2.24	1.37	1.25	2.14	1.34
m_{HH}	2.70	1.98	1.26	0.78	1.30	0.91	4.44	4.70	2.29	3.77	1.34	1.60	3.75	0.73
$\Delta\eta_{HH}$	6.95	1.37	1.52	1.07	1.51	1.63	1.99	0.98	1.38	1.94	1.16	3.90	1.19	0.99
$p_{T,H1}$	1.76	0.95	0.70	0.58	1.07	1.14	2.85	1.77	1.61	1.58	1.42	1.48	2.48	1.88
$p_{T,H2}$	2.81	1.83	2.83	1.17	1.82	1.06	0.92	1.49	2.64	2.50	1.74	1.51	2.24	1.00
η_{H1}	2.33	1.54	2.24	0.87	1.45	0.82	2.36	1.74	1.99	2.07	2.74	1.51	1.13	1.25
η_{H2}	2.62	1.35	4.93	1.11	0.85	0.75	1.33	1.16	1.37	1.36	2.07	1.17	1.26	1.16
$\Delta\phi_{HH}$	9.64	1.46	18.81	1.04	3.23	2.61	2.24	1.40	1.88	1.85	2.08	1.16	2.63	1.08
X_{Wt}	17.64	1.27	5.29	0.79	1.39	1.45	1.33	0.82	2.49	3.73	1.74	1.71	3.10	1.01
$p_{T,HH}$	6.06	1.13	1.76	1.00	1.50	1.23	1.80	0.80	2.54	0.93	2.35	0.87	2.00	0.79

Table 13.3: The χ^2 / (number of bins - 1) for the histograms after applying the X_{Wt} cut. In the top row, “3d discriminant” refers to the m_{HH} distribution with the $\Delta\eta_{HH}$ and X_{HH} categories.

Figure 13.14: m_{HH} for the background validation regions, shown after the X_{Wt} cut.

Outlook

This new paradigm for a robust background model is encouraging for the future of the $4b$ analysis for Run 3 and beyond. Below As the interpolation models the (m_{H1}, m_{H2}) dependence, this removes the need for the CR12 extrapolation systematic which is currently one of our limiting analysis uncertainties (Table 12.3). The validation regions studied above could provide an additional systematic for the interpolation method when propagating into an analysis. Another benefit of the interpolation is that we can draw an arbitrarily large number of samples to construct these background templates. This is important as the CMS $4b$ resolved analysis is currently limited by the statistical precision of their background templates [177]. Furthermore, alleviating the necessity of reweighting from a $2b$ region also helps the HL-LHC prospects where the higher rates make it impossible to collect a $2b$ sample without draconian trigger thresholds. Finally, a more robust background model might allow us to use ML-methods (i.e, NNs) for signal versus background discrimination to improve our sensitivity in the future. These avenues of exploration that open up promise to make the future of $HH \rightarrow 4b$ quite **bbbbb**right.

Conclusions

Journey before destination.

– Brandon Sanderson, *The Way of Kings*.

This thesis has presented a search for HH in the challenging $4b$ final state. As b -tagging is crucial for all HH searches, a part of this work included optimizing the latest recommendation of b -taggers, where the addition of the RNNIP low-level tagger, offered a 10% improvement to the high-level DL1r tagger. Furthermore, a new low-level b -tagger, DIPS was developed, which is four times faster to train than RNNIP, and further optimization shows a factor of two improvement compared to RNNIP.

We improved the $HH \rightarrow 4b$ analysis with an optimized event selection, physically motivated pairing algorithm, and sophisticated final discriminant. Furthermore, as the data-driven background estimate is the essential aspect of this analysis, a key innovation of this latest result is the array of background validation tests developed. As a result of these optimizations and robustness checks, our final observed (expected) limit on the HH SM signal strength is 5.4 (8.1), a 30% improvement compared to the luminosity scaling from the previous dataset. Propagating this result forward into the ATLAS combination gives an observed (expected) limit of the HH SM signal strength is 2.4 (2.9). $HH \rightarrow 4b$ additionally constrains the Higgs self-coupling in the kappa framework with observed (expected) limits of $[-3.9, 11.1]$ ($[-4.6, 10.8]$), while the observed (expected) constraints for a HH coupling to two vector bosons is $[-0.03, 2.11]$ ($[-0.05, 2.112]$). Results are additionally interpreted in both the SMEFT and HEFT frameworks.

As the quest for HH continues, so too do our opportunities for continued innovation, and to that end we presented a promising avenue for improving our background estimate. Incorporating these (and other) recent advancements from deep learning into the reconstruction and analysis allows us to maximize the physics potential of our growing datasets.

Life is a journey, physics is a process, and Kuhn postulates a part of progress in science involves

successively discarding old paradigms in lieu of new ones [178]. The past 60 years have shown the SM to be enormously successful at predicting a range of phenomena. It will take time to tell whether HH will be another confirmation of the SM paradigm, or herald deviations in the Higgs potential which could unlock answers to the open questions in physics. Either way, we'll have learned something that we hadn't known before, and this thesis analysis was a step forward along this path of exploration.

Appendix A

Additional FTAG optimizations

A.1 VR track jet optimizations

Hybrid sample modification

The track jet only uses charged constituents of the jet, so has a lower p_T distribution than the corresponding PFlow jets. Figure A.1 shows the correlation between the PFlow and corresponding VR track jet p_T s both in the $t\bar{t}$ and Z' samples, and a linear fit shows that on average the track jet p_T is approximately half of the jet p_T . The fitted slope is lower for the higher p_T track jets in the Z' sample since the VR jet clustering radius is inversely proportional to the jet p_T .

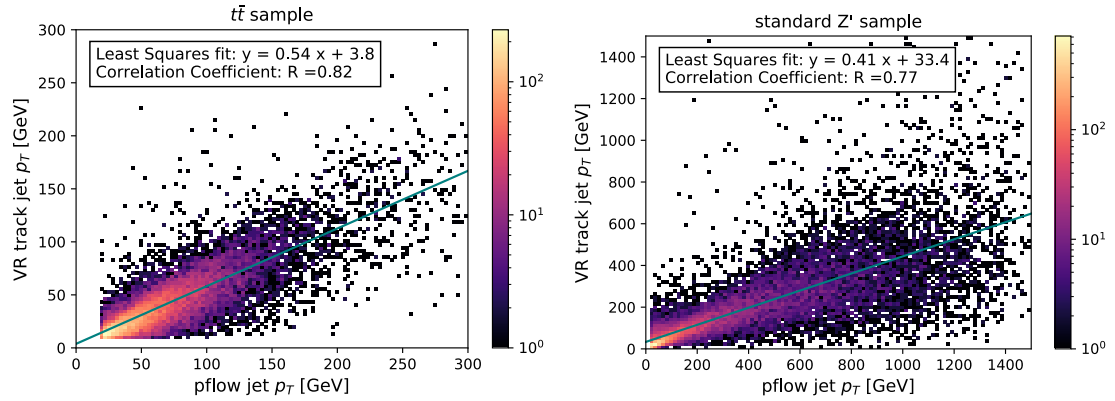


Figure A.1: Comparison of the PFlow and VR track jet p_T for jet reconstruction.

For the calo jets, the hybrid sample was created with a 250 GeV cut on the b -hadron of light/ c -jet p_T . To account for the difference in the track jet p_T spectrum, a 125 GeV jet p_T cut is used for light and c -jets to separate the $t\bar{t}$ and Z' samples. The same b -hadron 250 GeV cut is used for b -jets since this did not depend on the jet definition. The corresponding hybrid sample with the updated cut is shown in Figure A.2.

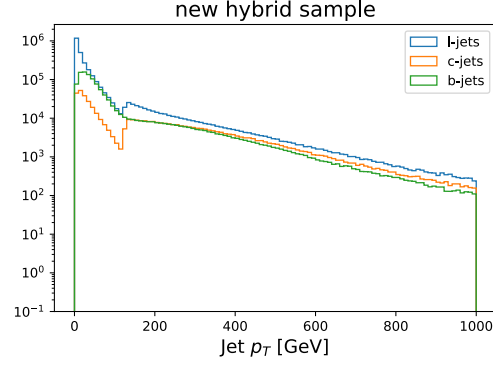


Figure A.2: The VR track jet p_T spectrum for the hybrid sample with the 125 GeV light and c -jet p_T cut.

Figure A.3 emphasizes how this updated cut improves our training performance. What is compared here are four different trainings on the VR track jet collection:

- training on $t\bar{t}$
- training on Z'
- training on a hybrid sample with a 250 GeV light and b -jet p_T cut
- training on a hybrid sample with a 150 GeV light and b -jet p_T cut : **our recommendation.**

With the hybrid definition with the 125 GeV cut, the hybrid training performs as well as the dedicated $t\bar{t}$ training for the $t\bar{t}$ evaluation, and performs as well as the dedicated Z' training for the Z' evaluation. This gave us confidence that the RNNIP was successfully optimized and ready for training DL1r.

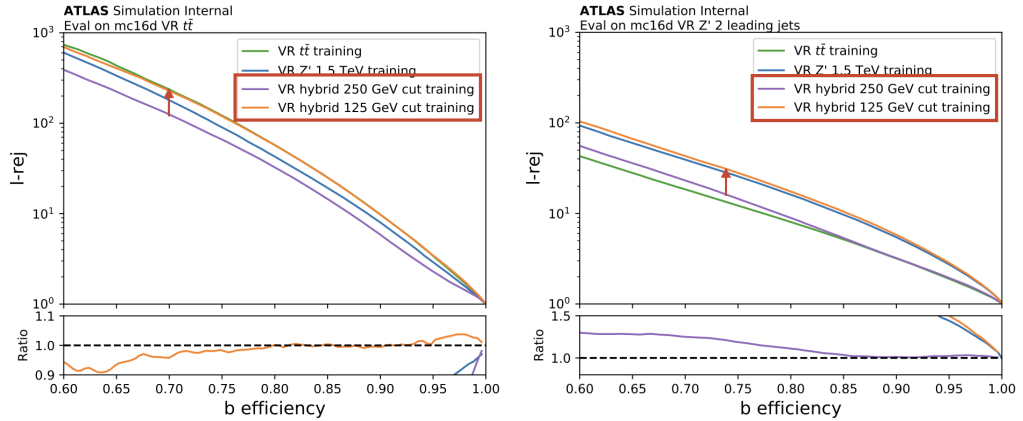


Figure A.3: Performance for trainings on four different samples for the VR track jet collection. The evaluations are done both on $t\bar{t}$ (left) and Z' sample (right). The red arrows on the plots emphasize the improved performance when the hybrid definition jet p_T cut was changed from 250 GeV to 125 GeV. The subpanels show the ratio of the performance with respect to the $t\bar{t}$ training.

How this improved on what we had before

Since this was the first time RNNIP was trained on the track jets, we compare the performance with what we had before. The calo-jet trained RNNIP networks had suboptimal performance on the VR track jet collection due to the domain shift of the underlying inputs. Figure A.4 compares three different trainings:

1. **EMTopo Rec:** The old recommendation trained on EMTopo – what was being applied before to VR track jets.
2. **Extended hybrid Pflow:** The new PFlow training (from Chapter 6.3.2), applied to VR track jets.
3. **New hybrid VR training:** NEW dedicated VR training.

The improvements from the VR track jets training improved the light jet rejection by more than a factor of 3 compared to what we had previously. The dedicated retraining also offered significant gains (in light rejection) compared to the PFlow training.

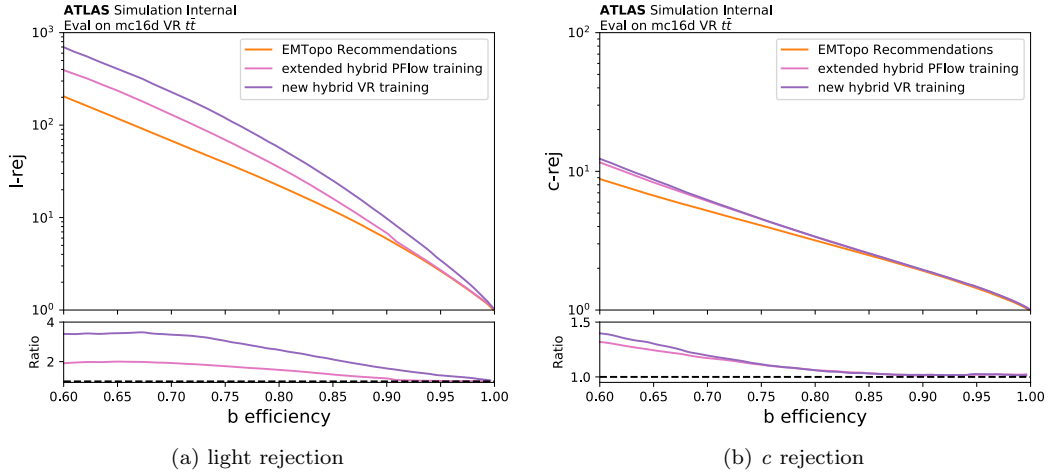


Figure A.4: Improvement in the performance from the dedicated VR track jet trainings (purple). The comparisons are to two previous calo-jet trainings: the old training EMTopo jets (orange) and the improved PFlow training (pink).

Where are these improvements coming from?

The RNN inputs (Table 6.3) with largest the differences from the PFlow inputs were the p_T^{frac} and the $\Delta R(trk, jet)$. These two variables are visualized below comparing the PFlow distributions (solid) with the VR track jets distributions (dashed). The $\Delta R(trk, jet)$ is lower because the track jet provides a better resolution for the direction of the initial quark (or gluon). Also, since the jet momentum is lower for track jets the p_T^{frac} is correspondingly higher. This motivated testing out the PFlow training, but rescaling $p_T^{frac} = \frac{p_T^{trk}}{p_T^{jet}} \rightarrow \frac{1}{2} \frac{p_T^{trk}}{p_T^{jet}}$. This modification recovers most of the performance of the $t\bar{t}$ VR training on the $t\bar{t}$ sample (left). The p_T^{frac} scaling also improves the Z' evaluation, but does not recover the performance of the dedicated VR Z' training.

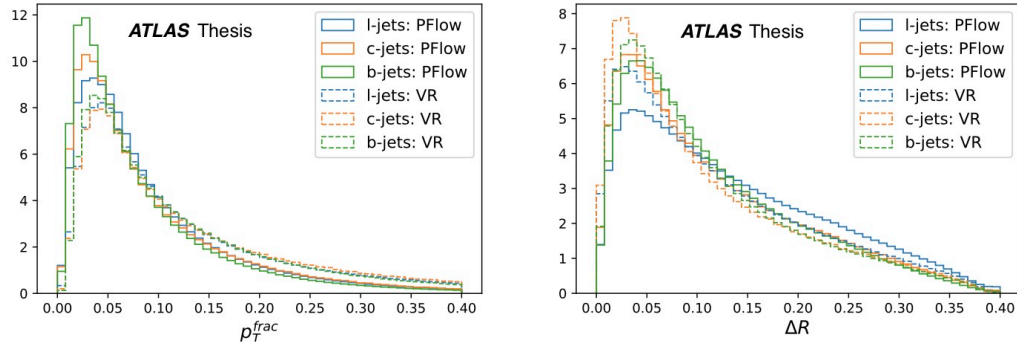
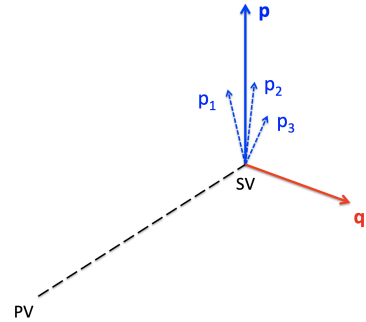


Figure A.5: $p_T^{frac} = p_T^{trk}/p_T^{jet}$ and the $\Delta R(trk, jet)$ in $t\bar{t}$ — the RNNIP inputs with the largest difference between the PFlow and VR track jet collections.

A.2 JetFitter mass constraint

The heavy hadron mass is especially crucial for b versus c -hadron identification, and including the conservation of momentum constraint to account for the neutral particles in the heavy-flavor hadron decay can help improve the mass resolution. This idea was first implemented by the SLD collaboration [179, 180], and this appendix motivates such a mass correction, and explains how it is encoded in the JetFitter algorithm. Let

- p : 4-vector of the **charged** particles
- q : 4-vector for **neutral** particles
- Let $p_{true} = p + q$ be the true 4-vector of charged and neutral particles.



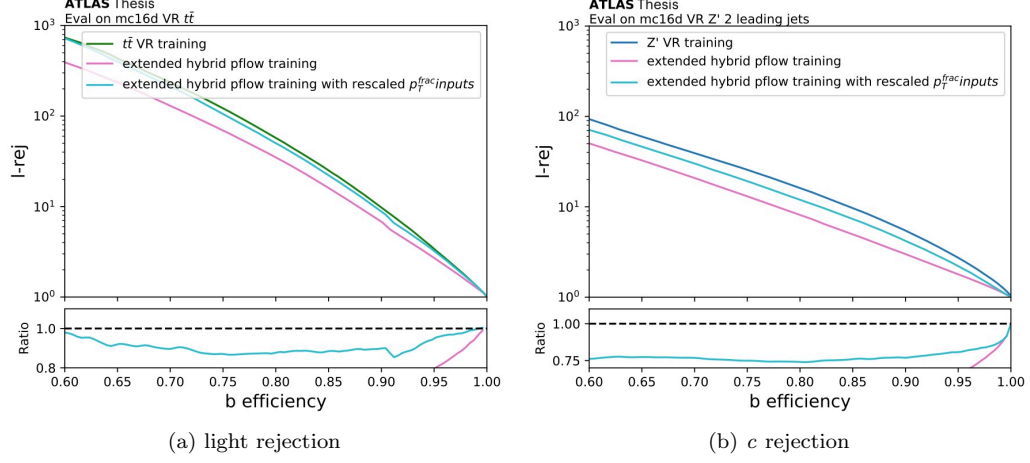


Figure A.6: Illustration of where the improvements of the dedicated PFlow training are coming from (for the light-jet rejection). The pink line is PFlow recommendation, while the cyan line is the same NN, but modifying the p_T^{frac} inputs $p_T^{frac} \Rightarrow \frac{1}{2}p_T^{frac}$. This is compared to a VR $t\bar{t}$ track jet training (left) and Z' track jet training (right).

Then $p = p_{true} - q$, which we can then square. We additionally make a simplifying assumption that the neutral particles are massless:

$$p^2 = p_{true}^2 - 2p \cdot q + q^2 \stackrel{0}{\rightarrow} \quad (\text{A.1})$$

Let $p^2 = m_{ch}^2$, the mass from the charged particles, and $p_{true}^2 = m^2$, the weakly decaying hadron mass. Then evaluate the right hand side of Eq. A.1 in the hadron's rest frame ($\vec{p}_{true}^{CM} = 0$). Then q is perpendicular to the hadron's flight axis $|\vec{q}^{CM}| = q_{\perp}$, where q_{\perp} is the component of \vec{q} perpendicular to the hadron's flight axis, which is invariant to boosts along the flight axis.

$$m_{ch}^2 = m^2 - 2 \left(m q_{\perp} - \vec{p}_{true}^{CM} \cdot \vec{q}^{CM} \right) \stackrel{0}{\rightarrow} \quad (\text{A.2})$$

$$m_{ch}^2 + p_{\perp}^2 = m^2 - 2mp_{\perp} + p_{\perp}^2 = (m - p_{\perp})^2 \quad (\text{A.3})$$

and solve for m :

$$m = \sqrt{m_{ch}^2 + p_{\perp}^2} + p_{\perp} \quad (\text{A.4})$$

where we took the $+$ solution of the $\sqrt{}$ for the physical solution of the positive hadron mass.

Figure A.7 shows the impact of this correction at truth level comparing no correction (red), the correction from Eq. A.4 (blue) and two different cuts to constrain the high mass tails from becoming too large: restricting $M < 2M_{ch}$ (green) and $p_{\perp} < M_{ch}$ (purple). All of the correction formulas

significantly increase the reconstructed b -hadron mass, and therefore improve the b versus c -hadron separation power.

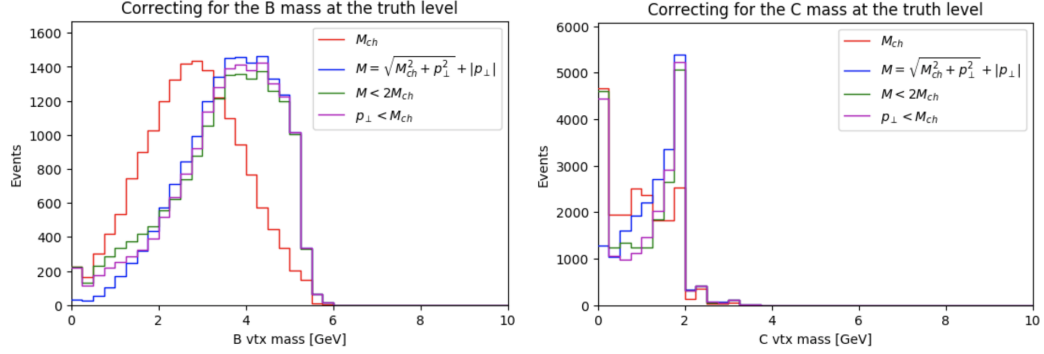


Figure A.7: Reconstruction of the B -hadron (left) and D -hadron (right) masses from the truth charged particles.

Given this truth-level intuition, we next show the correction formula in the JetFitter algorithm.

How JF applies the mass correction:

- It uses the scalar sum of the tracks, $p_{\perp} = \sum_i p_{\perp}^{(i)}$, where i runs over the tracks in the jet. Then the vector sum over the tracks is replaced with the scalar sum over the tracks in the formula as: $m = \sqrt{m_{ch}^2 + p_{\perp}^2} + |p_{\perp}|$.
- To constrain the tails, if $m > 5$, $m \leftarrow 5 [1 + 2 \arctan(\pi(m - 5))]$.

Although the intuition for using the scalar sum of the track p_T s is not motivated by the conservation of momentum argument above, Figure A.8 shows that using the scalar sum over the tracks for the mass correction gave a lower background mistag rate just computing the discrimination from the JetFitter reconstructed mass.

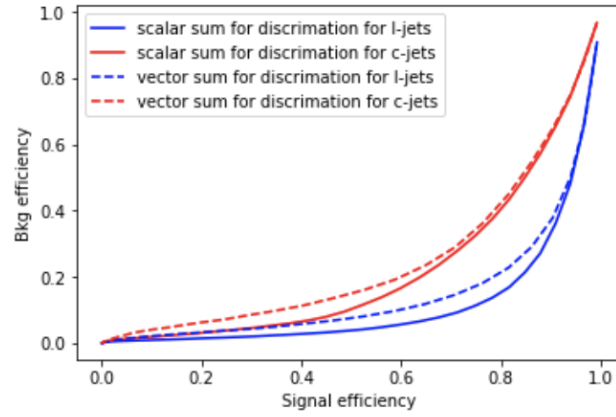


Figure A.8: Comparison of the ROC curves from the JF corrected mass whether the scalar sum for p_{\perp} is used (solid) or the vector sum for p_{\perp} is used (dashed).

The motivation for why the scalar sum does better is shown in Figure A.9. Using the truth charged particles, the reconstructed masses are compared as one or more of the truth particles are randomly dropped. As more tracks are lost, the scalar sum does increasingly better over the vector sum. On average, JetFitter fails to find 2.15 of the truth particles, so using this form for mass correction helps compensate for the limitations in the track selection efficiency.

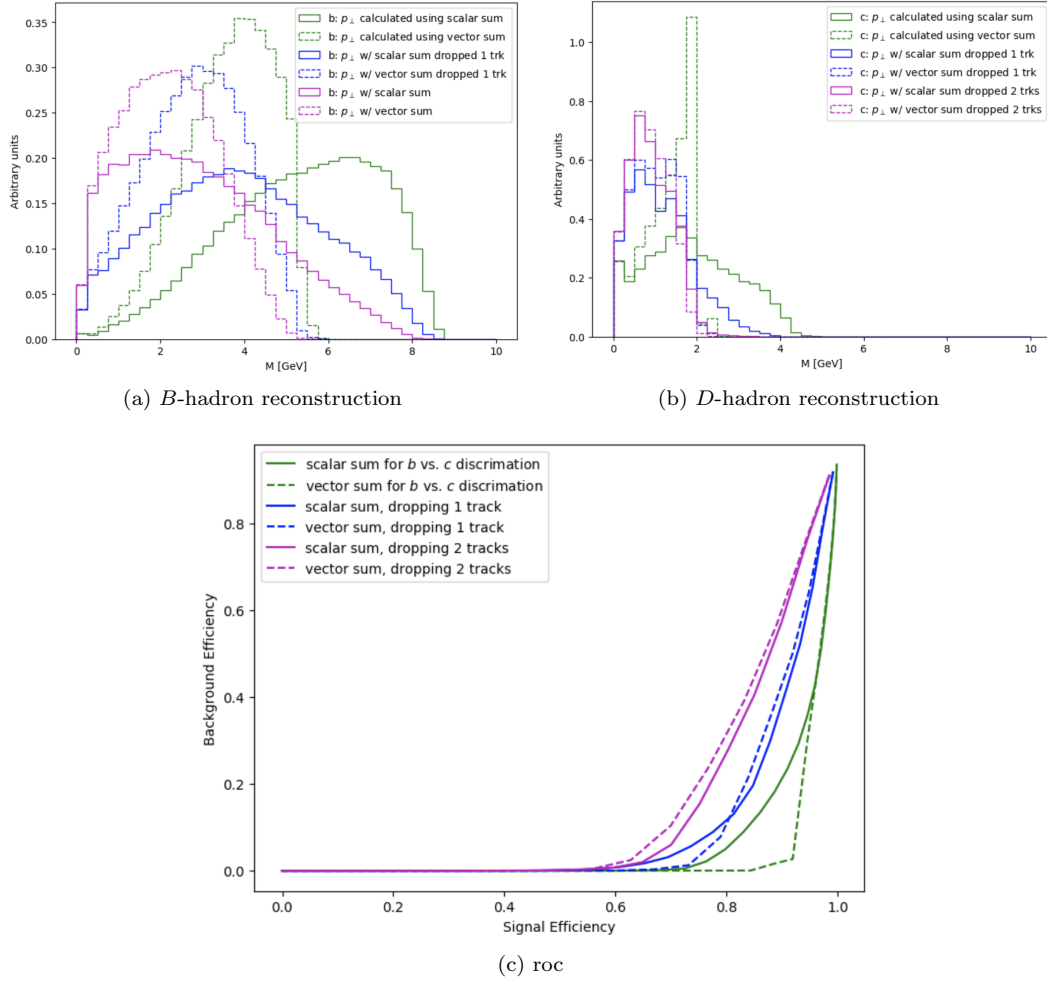


Figure A.9: Reconstruction of the B (left) and D (right) hadron masses from the truth charged particles. The solid lines show the masses reconstructed from the scalar sum, while the dashed lines show the reconstruction from the vector sum. The purple (blue) lines show the result for randomly dropping one (two) of the truth particle(s). The roc curve demonstrates that as we drop more tracks,

Appendix B

Further statistics details

B.1 Asymptotics approximation

As given by [132], Wald's theorem allows us to write our test statistic as:

$$q_\mu = \begin{cases} \frac{(\mu - \mu')^2}{\sigma^2}, \\ 0 \end{cases} \quad \hat{\mu} > \mu.$$

$$f(q_\mu | \mu') = \Phi \left(\frac{(\mu - \mu')^2}{\sigma^2} \right) \delta(q_\mu) + \frac{1}{2} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{q_\mu}} \exp \left[-\frac{1}{2} \left(\sqrt{q_\mu} - \frac{(\mu - \mu')^2}{\sigma^2} \right)^2 \right] \quad (\text{B.1})$$

The Asimov data sets the histograms \vec{n} and \vec{m} to their expectation values:

$$n_{i,A} = \mu' s_i(\theta) + b_i(\theta) \quad (\text{B.2})$$

$$m_{i,A} = u_i(\theta). \quad (\text{B.3})$$

This Asimov dataset is critical for finding the σ^2 that describes the distribution over $\hat{\mu}$. When the likelihood is evaluated on the Asimov test statistic gives the Asimov test-statistic $q_{\mu,A}$.

$$q_{\mu,A} = \frac{(\mu - \mu')^2}{\sigma_A^2} = \implies \sigma_A^2 = \frac{(\mu - \mu')^2}{q_{\mu,A}}$$

To compare the pdfs $f(q_\mu | \mu')$ reconstructed from samples and the result from the asymptotic formula in Eq. B.1, consider a cut-and-count analysis where the signal region (SR) has s signal events, and the background yield is specified by a nuisance parameter (NP) b expected background with b events. Let the number of events that we observe in the SR be n , so then the expected value for the observed event yield is: $\mathbb{E}[n] = \mu s + b$. We additionally consider that we have a measurement in a

control region where m events are observed, and this control region where the background yield is scaled by τ relative to the SR, i.e., $\mathbb{E}[m] = \tau b$.

Since this is now just a single bin counting experiment, the likelihood from Eq 8.1 simplifies to the form:

$$\mathcal{L}(\mu, b) = \frac{(\mu s + b)^n}{n!} e^{-\mu s + b} \frac{(\mu s + b)^n}{n!} e^{-\mu s + b} \quad (\text{B.4})$$

Figure ?? compares the analytic asymptotic formula for the test statistic's distribution to the result from throwing toys. For the model with $s = 6$, $b = 9$, and $\tau = 1$, the red histogram shows samples for $f(q_1|0)$ while the blue histogram shows samples for $f(q_1|1)$. The result from the asymptotic formula are shown in the solid black lines and are in good agreement with the pdfs reconstructed from the test statistic samples. Since there is a signal in the generated dataset, the samples from the $\mu' = 0$ hypothesis ($f(q_1|0)$ in red) have larger q_1 s since this is not compatible with the data. the median of the $f(q_1|0)$ distribution is the $q_{1,A}$ in the dotted line.

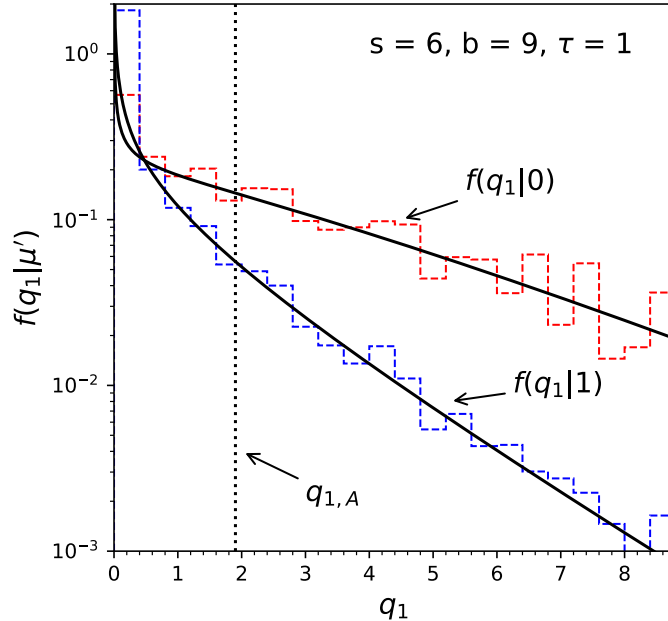


Figure B.1: Comparison of the pdfs reconstructed from test statistic samples (red and blue dashed histograms) and the result from the asymptotic formula (solid black lines) [132]

B.2 A pedagogical motivation for the CLs test statistic

There can be a problem for the frequentist interpretation if the number of observed events is less than the number of predicted events (in our notation, $n_i < b_i$). Consider a Poisson experiment where

$b = 8$ background events are expected, but only 3 events are observed (example from [181]). For the model $n = s + b$, we can calculate the 95% upper limit on the true expected number of events by summing the poisson probabilities of observing this many events or fewer (which is what is also defined as CL_{s+b}):

$$Prob(\text{observing} \leq 3 \text{ events}) = CL_{s+b} = \sum_{r=0}^3 \frac{n^r e^{-n}}{r!}. \quad (\text{B.5})$$

Figure B.2 (blue line) shows the CL_{s+b} for a range of possible n . Setting $CL_{s+b} = 0.05$ and solving for n using the transcendental equation $\sum_{r=0}^3 \frac{7.75^r e^{-7.75}}{r!} = 0.05$ gives $n = 7.75$. Then the upper limit on the signal yield is $s = n - b = n - 8 = -0.25$, an unphysical result.

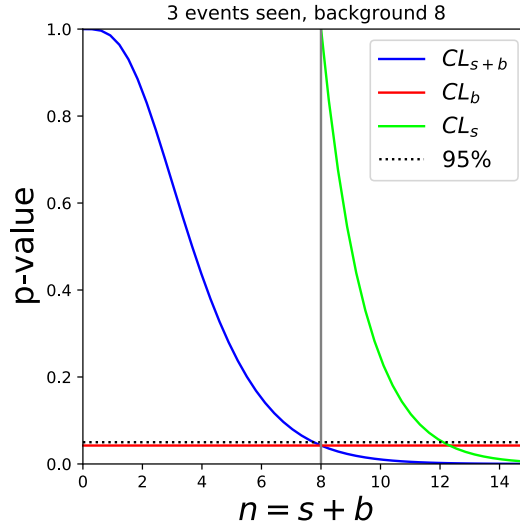


Figure B.2: A comparison of the definitions for CL_{s+b} , CL_b , and $CL_s = CL_{s+b}/CL_b$ associated to a model $n = s + b$ [181].

To avoid the pathologies excluding negative signal strengths, we define the 95% upper limit as $CL_s = 0.05$, where

$$CL_s = \frac{CL_{s+b}}{CL_b}. \quad (\text{B.6})$$

- CL_{s+b} : Probability of having this many events (or more) given an expected yield of $s + b$.
- CL_b : Probability of having this many events (or fewer) for an expected yield of b (and no signal).

Appendix C

Extra analysis plots

This chapter includes some extra analysis definitions and validation plots which although not necessary for the main storyline of the $4b$ analysis description, are useful for validating the performance, or showing the robustness across the various configurations.

C.1 Event selection

To complement the plots from Figure 10.12, pairing accuracy plots for all the events passing the $4p_T \geq 40$ GeV cut (i.e, no requirement that the correct pair exists) are shown in Figure C.1

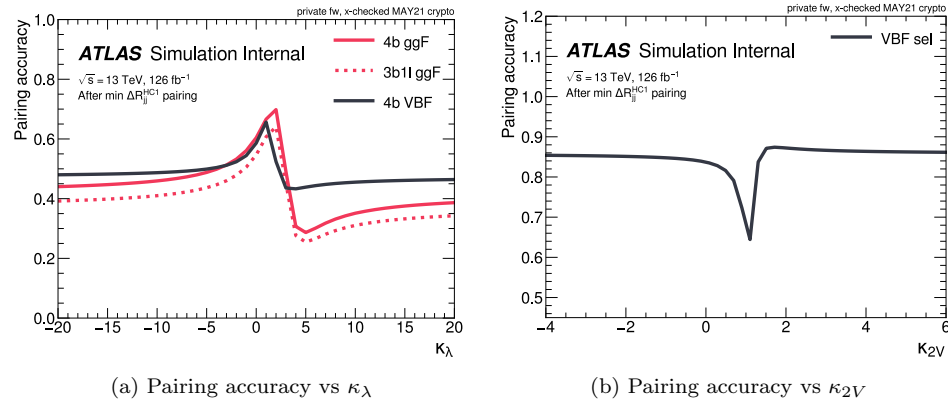


Figure C.1: The pairing accuracy as a function of κ_λ and κ_{2V}

Figure C.2 shows the pairing accuracy as a function of truth m_{HH} for a few different coupling values for when we have selected the correct jets. It demonstrates that the pairing accuracy is mostly determined on the truth m_{HH} and there's only a mild dependence on the coupling values for the VBF selection.

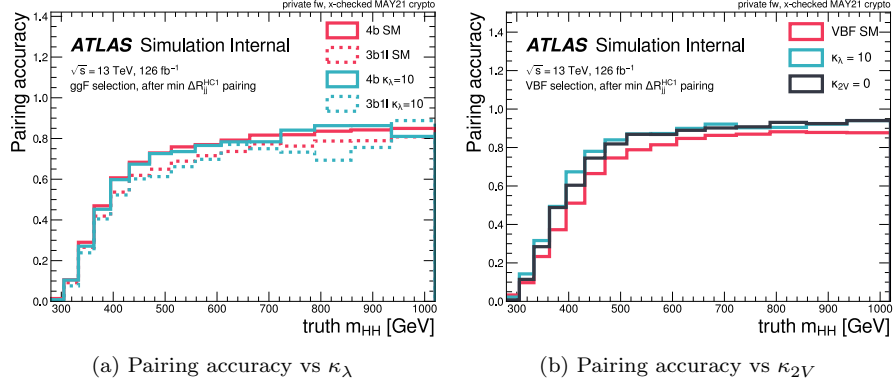
(a) Pairing accuracy vs κ_λ (b) Pairing accuracy vs κ_{2V} Figure C.2: The pairing accuracy as a function of κ_λ and κ_{2V}

Figure C.3 visualizes the impact of the X_{Wt} cut in the $4b$ and $2b$ massplanes for VBF.

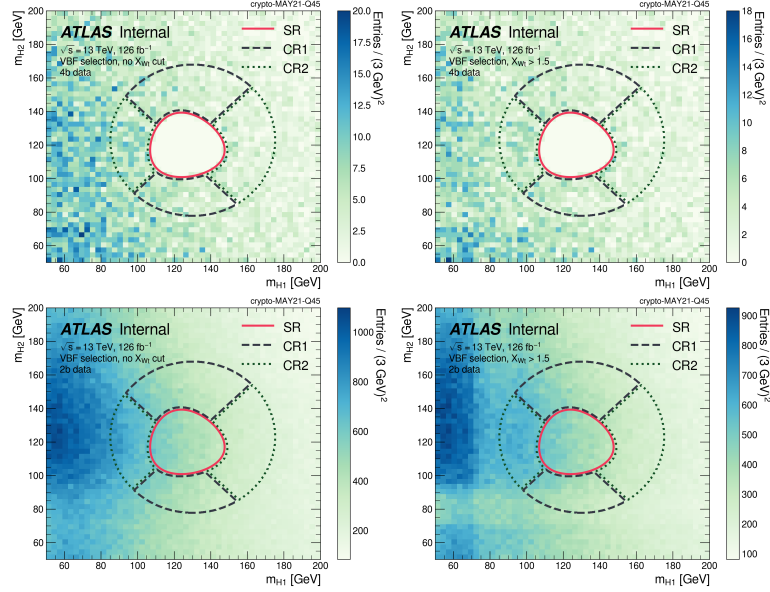


Figure C.3: Impact of the X_{wt} cut for the VBF HC massplanes. The top row shows the $4b$ distributions before (left) and after (right) the X_{wt} cut, and the bottom row shows the same for $2b$.

C.2 Cutflows

ggF HH MC simulation (normalized to 126.1 fb^{-1}):

- Table C.1: 4b ggF channel events for the SM and $\kappa_\lambda = 10$ ggF HH signals.
- Table C.2: 4b VBF channel events for the SM (a) and $\kappa_\lambda = 10$ (b) ggF HH signals.

VBF HH MC (normalized to 126.1 fb^{-1}):

- Table C.3: 4b VBF channel events for the SM, $\kappa_\lambda = 10$ and $\kappa_{2V} = 0$ VBF HH signals.
- Table C.4: 4b ggF channel events for the SM, $\kappa_\lambda = 10$, and $\kappa_{2V} = 0$ VBF HH signals.

$t\bar{t}$ MC simulation (normalized to 126.1 fb^{-1}):

- Table C.5: 4b ggF channel events for semi-leptonic and all hadronic $t\bar{t}$ MC.
- Table C.6: 4b VBF channel events for semi-leptonic and all hadronic $t\bar{t}$ MC.

Table C.1: **ggF HH MC** yields at each step in the analysis event selection for 4b events in the **ggF channel** normalized to 126.1fb^{-1} , alongside the ratio of each yield to the initial yield and to the yield for the previous cut.

(a) 4b SM ggF HH MC simulation (ggF channel)			
	Yield	Yield / Pre-selection	Yield / Prior cut
Pass NTuple Preselection	526.6	1	-
Trigger	475	0.9019	0.9019
Trigger Buckets	419	0.7956	0.8821
Multiply FTAG, trig, + JVT SFs	381.8	0.725	0.9112
ggF channel	376.6	0.7151	0.9864
≥ 4 central jets, ≥ 2 b -tags	322.4	0.6122	0.8561
≥ 4 b -tags	86	0.1633	0.2668
$ \Delta\eta_{hh} < 1.5$	71.85	0.1364	0.8355
Top Veto	60.4	0.1147	0.8406
Signal Region	29.1	0.05525	0.4817
Control Region 2	7.137	0.01355	0.1182
Control Region 1	11.41	0.02166	0.1889

(b) 4b $\kappa_\lambda = 10$ ggF HH MC simulation (ggF channel)			
	Yield	Yield / Pre-selection	Yield / Prior cut
Pass NTuple Preselection	7338	1	-
Trigger	6917	0.9427	0.9427
Trigger Buckets	6378	0.8692	0.922
Multiply FTAG, trig, + JVT SFs	5279	0.7194	0.8278
ggF channel	5198	0.7084	0.9846
≥ 4 central jets, ≥ 2 b -tags	4314	0.588	0.83
≥ 4 b -tags	1002	0.1365	0.2322
$ \Delta\eta_{hh} < 1.5$	850.6	0.1159	0.8492
Top Veto	569	0.07754	0.6689
Signal Region	182.7	0.0249	0.3211
Control Region 2	66.06	0.009003	0.1161
Control Region 1	86.18	0.01175	0.1515

Table C.2: **ggF HH MC** yields at each step in the analysis event selection for 4b events in the **VBF channel** normalized to 126.1fb^{-1} , alongside the ratio of each yield to the initial yield and to the yield for the previous cut.

(a) 4b SM ggF HH MC simulation (VBF channel)			
	Yield	Yield / Pre-selection	Yield / Prior cut
Pass NTuple Preselection	526.6	1	-
Trigger	475	0.9019	0.9019
Trigger Buckets	419	0.7956	0.8821
Multiply FTAG, trig, + JVT SFs	381.8	0.725	0.9112
VBF channel	5.208	0.00989	0.01364
≥ 4 central jets, ≥ 2 b -tags	5.155	0.009789	0.9898
≥ 4 b -tags	1.14	0.002165	0.2212
$ \Delta\eta_{hh} < 1.5$	1.14	0.002165	1
Top Veto	1.008	0.001914	0.8838
Signal Region	0.4833	0.0009177	0.4796
Control Region 2	0.123	0.0002336	0.122
Control Region 1	0.1703	0.0003234	0.169

(b) 4b $\kappa_\lambda = 10$ ggF HH MC simulation (VBF channel)			
	Yield	Yield / Pre-selection	Yield / Prior cut
Pass NTuple Preselection	7338	1	-
Trigger	6917	0.9427	0.9427
Trigger Buckets	6378	0.8692	0.922
Multiply FTAG, trig, + JVT SFs	5279	0.7194	0.8278
VBF channel	81.14	0.01106	0.01537
≥ 4 central jets, ≥ 2 b -tags	80.22	0.01093	0.9888
≥ 4 b -tags	15.29	0.002084	0.1906
$ \Delta\eta_{hh} < 1.5$	15.29	0.002084	1
Top Veto	11.15	0.001519	0.729
Signal Region	3.099	0.0004223	0.278
Control Region 2	1.449	0.0001975	0.13
Control Region 1	1.851	0.0002522	0.166

Table C.3: **VBF HH MC** yields at each step in the analysis event selection for 4b events in the **VBF channel** normalized to 126.1fb^{-1} , alongside the ratio of each yield to the initial yield and to the yield for the previous cut.

(a) 4b SM VBF HH MC simulation (VBF channel)			
	Yield	Yield / Pre-selection	Yield / Prior cut
Pass NTuple Preselection	22.26	1	-
Trigger	20.67	0.9287	0.9287
Trigger Buckets	18.39	0.8261	0.8895
Multiply FTAG, trig, + JVT SFs	16.14	0.7251	0.8778
VBF channel	2.238	0.1005	0.1387
≥ 4 central jets, ≥ 2 b -tags	2.223	0.09986	0.9931
≥ 4 b -tags	0.7449	0.03347	0.3351
$ \Delta\eta_{hh} < 1.5$	0.7449	0.03347	1
Top Veto	0.6722	0.0302	0.9024
Signal Region	0.3265	0.01467	0.4857
Control Region 2	0.09092	0.004085	0.1352
Control Region 1	0.1149	0.005164	0.171

(b) 4b $\kappa_\lambda = 10$ VBF HH MC simulation (VBF channel)			
	Yield	Yield / Pre-selection	Yield / Prior cut
Pass NTuple Preselection	1573	1	-
Trigger	1465	0.9317	0.9317
Trigger Buckets	1303	0.8289	0.8896
Multiply FTAG, trig, + JVT SFs	1092	0.6941	0.8374
VBF channel	136.1	0.08651	0.1246
≥ 4 central jets, ≥ 2 b -tags	135.1	0.08589	0.9929
≥ 4 b -tags	46.14	0.02934	0.3416
$ \Delta\eta_{hh} < 1.5$	46.14	0.02934	1
Top Veto	34.84	0.02216	0.7551
Signal Region	14.24	0.009055	0.4087
Control Region 2	4.23	0.00269	0.1214
Control Region 1	5.03	0.003198	0.1443

(c) 4b $\kappa_{2V} = 0$ VBF HH MC simulation (VBF channel)			
	Yield	Yield / Pre-selection	Yield / Prior cut
Initial (Unweighted for MC)	1.331e+06	-	-
Pass NTuple Preselection	626.1	1	-
Trigger	513.5	0.8201	0.8201
Trigger Buckets	444.6	0.71	0.8658
Multiply FTAG, trig, + JVT SFs	405.2	0.6471	0.9114
VBF channel	70.74	0.113	0.1746
≥ 4 central jets, ≥ 2 b -tags	70.3	0.1123	0.9939
≥ 4 b -tags	27.63	0.04412	0.393
$ \Delta\eta_{hh} < 1.5$	27.63	0.04412	1
Top Veto	26.52	0.04236	0.9601
Signal Region	17.29	0.02761	0.6517
Control Region 2	3.074	0.00491	0.1159
Control Region 1	3.805	0.006078	0.1435

Table C.4: **VBF HH MC** yields at each step in the analysis event selection for 4b events in the **ggF channel** normalized to 126.1fb^{-1} , alongside the ratio of each yield to the initial yield and to the yield for the previous cut.

(a) 4b SM VBF HH MC simulation (ggF channel)			
	Yield	Yield / Pre-selection	Yield / Prior cut
Pass NTuple Preselection	22.26	1	-
Trigger	20.67	0.9287	0.9287
Trigger Buckets	18.39	0.8261	0.8895
Multiply FTAG, trig, + JVT SFs	16.14	0.7251	0.8778
ggF channel	13.9	0.6246	0.8613
≥ 4 central jets, ≥ 2 b -tags	10.75	0.4829	0.7731
≥ 4 b -tags	1.87	0.08402	0.174
$ \Delta\eta_{hh} < 1.5$	0.9419	0.04232	0.5036
Top Veto	0.736	0.03307	0.7814
Signal Region	0.2352	0.01057	0.3195
Control Region 2	0.07797	0.003503	0.1059
Control Region 1	0.1094	0.004914	0.1486

(b) 4b $\kappa_\lambda = 10$ VBF HH MC simulation (ggF channel)			
	Yield	Yield / Pre-selection	Yield / Prior cut
Pass NTuple Preselection	1573	1	-
Trigger	1465	0.9317	0.9317
Trigger Buckets	1303	0.8289	0.8896
Multiply FTAG, trig, + JVT SFs	1092	0.6941	0.8374
ggF channel	955.5	0.6076	0.8754
≥ 4 central jets, ≥ 2 b -tags	756.5	0.481	0.7917
≥ 4 b -tags	134.5	0.08551	0.1778
$ \Delta\eta_{hh} < 1.5$	109.4	0.06954	0.8132
Top Veto	76.41	0.04859	0.6988
Signal Region	25.19	0.01602	0.3297
Control Region 2	8.892	0.005654	0.1164
Control Region 1	11.42	0.007262	0.1495

(c) 4b $\kappa_{2V} = 0$ VBF HH MC simulation (ggF channel)			
	Yield	Yield / Pre-selection	Yield / Prior cut
Initial (Unweighted for MC)	1.331e+06	-	-
Pass NTuple Preselection	626.1	1	-
Trigger	513.5	0.8201	0.8201
Trigger Buckets	444.6	0.71	0.8658
Multiply FTAG, trig, + JVT SFs	405.2	0.6471	0.9114
ggF channel	334.4	0.5341	0.8254
≥ 4 central jets, ≥ 2 b -tags	260.5	0.416	0.7788
≥ 4 b -tags	65.23	0.1042	0.2504
$ \Delta\eta_{hh} < 1.5$	46.37	0.07405	0.7108
Top Veto	43.1	0.06884	0.9296
Signal Region	22.97	0.03669	0.533
Control Region 2	4.854	0.007752	0.1126
Control Region 1	8.26	0.01319	0.1916

Table C.5: $t\bar{t}$ MC yields at each step in the analysis event selection for 4b events in the **ggF channel** normalized to 126.1fb^{-1} , alongside the ratio of each yield to the initial yield and to the yield for the previous cut.

(a) 4b Non-all hadronic $t\bar{t}$ MC simulation (ggF channel)			
	Yield	Yield / Pre-selection	Yield / Prior cut
Pass NTuple Preselection	6.698e+06	1	-
Trigger	6.215e+06	0.9279	0.9279
Trigger Buckets	5.542e+06	0.8274	0.8917
ggF channel	5.488e+06	0.8193	0.9903
≥ 4 central jets, ≥ 2 b -tags	4.313e+06	0.6439	0.7859
≥ 4 b -tags	3.424e+04	0.005112	0.007939
$ \Delta\eta_{hh} < 1.5$	1.989e+04	0.002969	0.5807
Top Veto	1.07e+04	0.001598	0.5383
Signal Region	403.5	6.024e-05	0.0377
Control Region 2	642.3	9.59e-05	0.06
Control Region 1	734.8	0.0001097	0.06864

(b) 4b All hadronic $t\bar{t}$ MC simulation (ggF channel)			
	Yield	Yield / Pre-selection	Yield / Prior cut
Pass NTuple Preselection	6.698e+06	1	-
Trigger	6.215e+06	0.9279	0.9279
Trigger Buckets	5.542e+06	0.8274	0.8917
ggF channel	5.488e+06	0.8193	0.9903
≥ 4 central jets, ≥ 2 b -tags	4.313e+06	0.6439	0.7859
≥ 4 b -tags	3.424e+04	0.005112	0.007939
$ \Delta\eta_{hh} < 1.5$	1.989e+04	0.002969	0.5807
Top Veto	1.07e+04	0.001598	0.5383
Signal Region	403.5	6.024e-05	0.0377
Control Region 2	642.3	9.59e-05	0.06
Control Region 1	734.8	0.0001097	0.06864

Table C.6: $t\bar{t}$ MC yields at each step in the analysis event selection for 4b events in the **VBF channel** normalized to 126.1fb^{-1} , alongside the ratio of each yield to the initial yield and to the yield for the previous cut.

(a) 4b Semi-leptonic $t\bar{t}$ MC simulation (VBF channel)			
	Yield	Yield / Pre-selection	Yield / Prior cut
Pass NTuple Preselection	6.698e+06	1	-
Trigger	6.215e+06	0.9279	0.9279
Trigger Buckets	5.542e+06	0.8274	0.8917
VBF channel	5.383e+04	0.008036	0.009713
≥ 4 central jets, ≥ 2 b -tags	5.25e+04	0.007839	0.9754
≥ 4 b -tags	351.8	5.252e-05	0.0067
$ \Delta\eta_{hh} < 1.5$	351.8	5.252e-05	1
Top Veto	219.2	3.272e-05	0.623
Signal Region	8.882	1.326e-06	0.04053
Control Region 2	13.06	1.95e-06	0.05959
Control Region 1	13.07	1.951e-06	0.05964

(b) 4b All hadronic $t\bar{t}$ MC simulation (VBF channel)			
	Yield	Yield / Pre-selection	Yield / Prior cut
Pass NTuple Preselection	6.698e+06	1	-
Trigger	6.215e+06	0.9279	0.9279
Trigger Buckets	5.542e+06	0.8274	0.8917
VBF channel	5.383e+04	0.008036	0.009713
≥ 4 central jets, ≥ 2 b -tags	5.25e+04	0.007839	0.9754
≥ 4 b -tags	351.8	5.252e-05	0.0067
$ \Delta\eta_{hh} < 1.5$	351.8	5.252e-05	1
Top Veto	219.2	3.272e-05	0.623
Signal Region	8.882	1.326e-06	0.04053
Control Region 2	13.06	1.95e-06	0.05959
Control Region 1	13.07	1.951e-06	0.05964

C.3 Background QCD composition

# of true b jets	QCD background [%]		
	2016	2017	2018
≥ 5	4.4	4.0	2.3
4	67.5	75.0	69.0
3	11.7	12.4	14.5
2	16.0	8.4	14.1
≤ 1	0.4.	0.1	0.1

Table C.7: Background flavor composition of the # of b -jets in the 4b SR in pythia QCD sample.

C.4 Background systematics plots

For completeness, we also show the templates by category with the different quadrant contributions overlaid in Figure C.4.

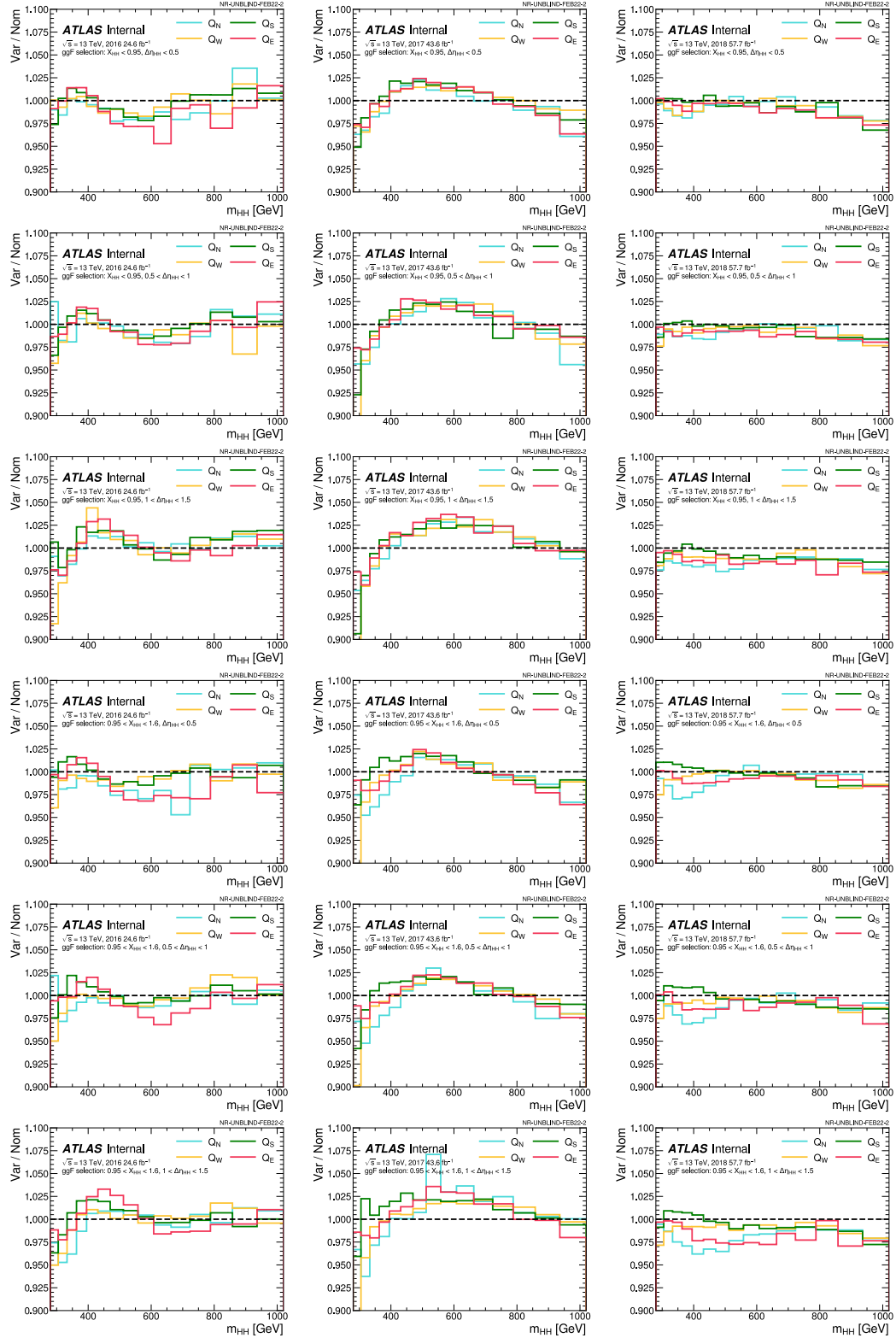


Figure C.4: Impact of background shape nuisance parameter variation on m_{HH} in different kinematic categories for the ggF channel. Each column is a different year of the ggF channel templates while the rows show each category with the SR NP quadrants overlaid.

C.5 Pre-fit plots

The distributions for the prefit background model compared to the 4b SR data are shown in Figure C.5 and Figure C.6, respectively.

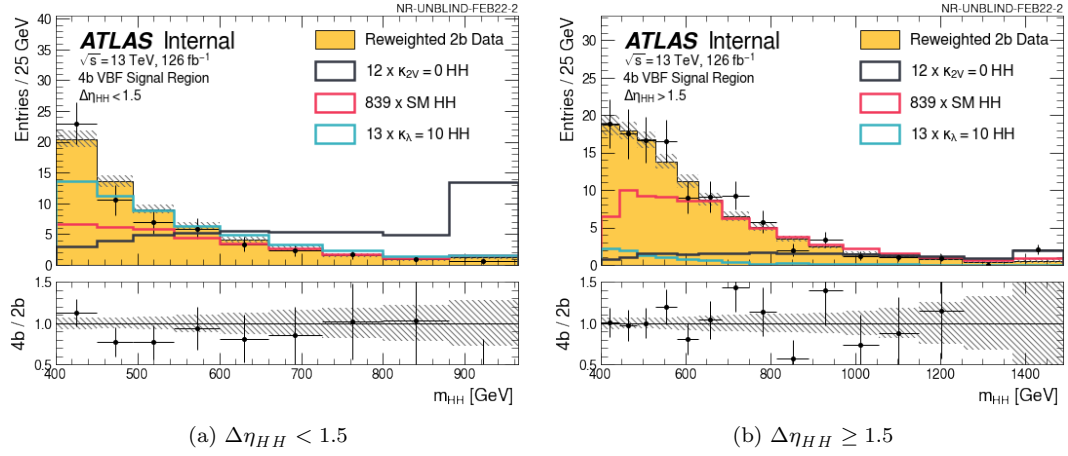


Figure C.5: 4b VBF background and selected signal histograms for 2016, 2017, and 2018 with the proposed binning and categorization. To visualize these signals they are scaled by $\alpha = 839$ and 13, and 12 for the VBF SM, $\kappa_\lambda = 10$, and $\kappa_{2V}=0$ signals, respectively.

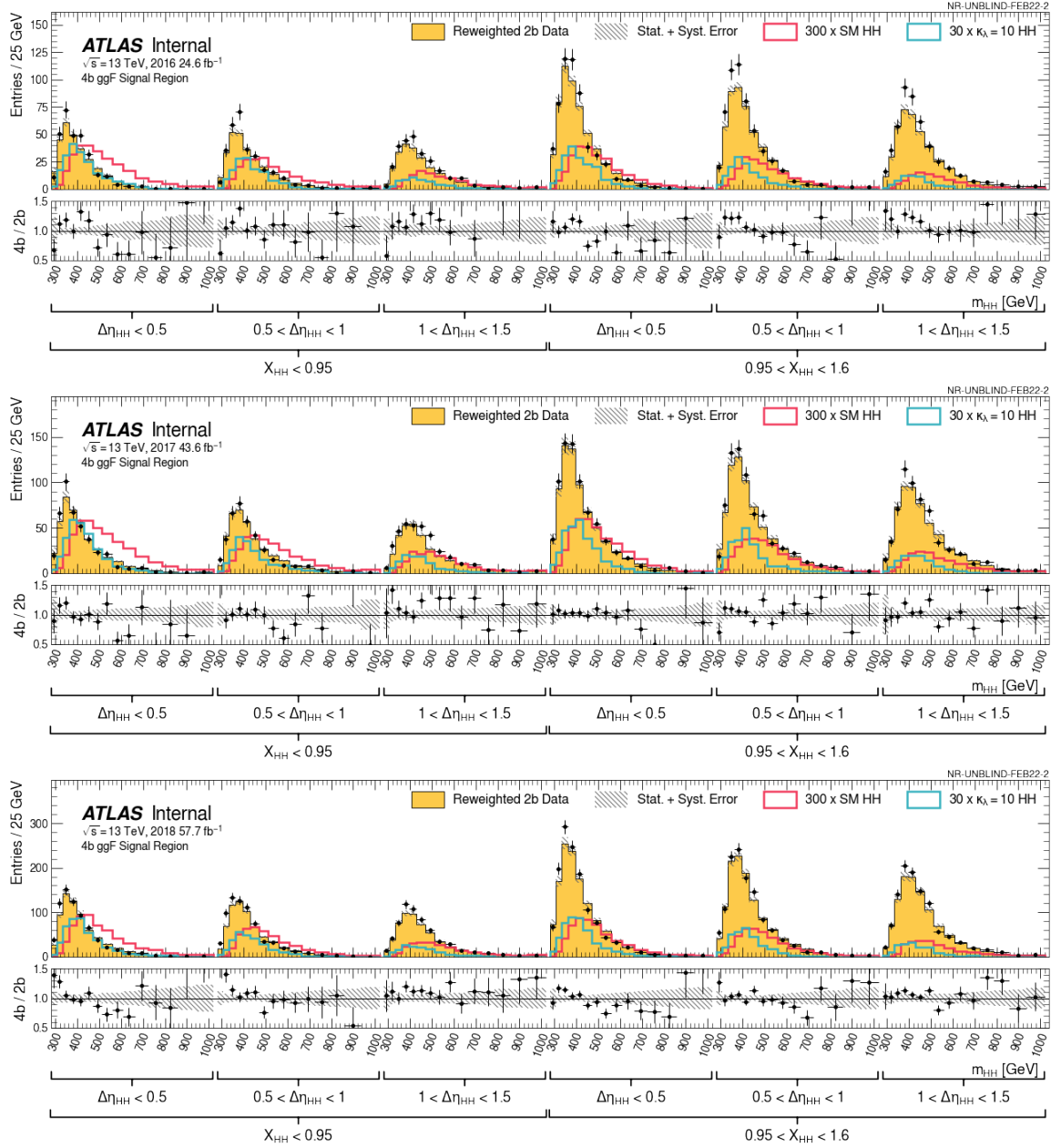


Figure C.6: 4b ggF background and selected signal histograms for 2016 (top), 2017 (middle), and 2018 (bottom) with the categorization. To visualize the signals they are scaled by $\alpha = 300$ and 30 for the ggF SM NR and $\kappa_\lambda=10$ signals, respectively.

C.6 Additional limits

Table C.8 shows the limits on the SM signal strength with the impact of the ggF and VBF analysis channels, and including the ggF and VBF signal production modes.

Figure C.7 shows the combined κ_λ and κ_V limits for the three main analysis channels. The $4b$ channel isn't very sensitive in the κ_λ scan, but it drives the κ_V limit [163].

Figure C.8 shows a comparison between the signal and backgrounds for the three most sensitive HH channels: $bb\gamma\gamma$, $bb\tau\tau$, and $4b$. The signal, background, S/B and (quadrature summed) S/\sqrt{B} the most sensitive HH channels: $bb\gamma\gamma$, $bb\tau\tau$, and $4b$ [7, 182, 183]. For this heuristic comparison, a loose cut on the $bb\tau\tau$ MVA discriminants was used to mimic the $bb\gamma\gamma$ BDT categories.

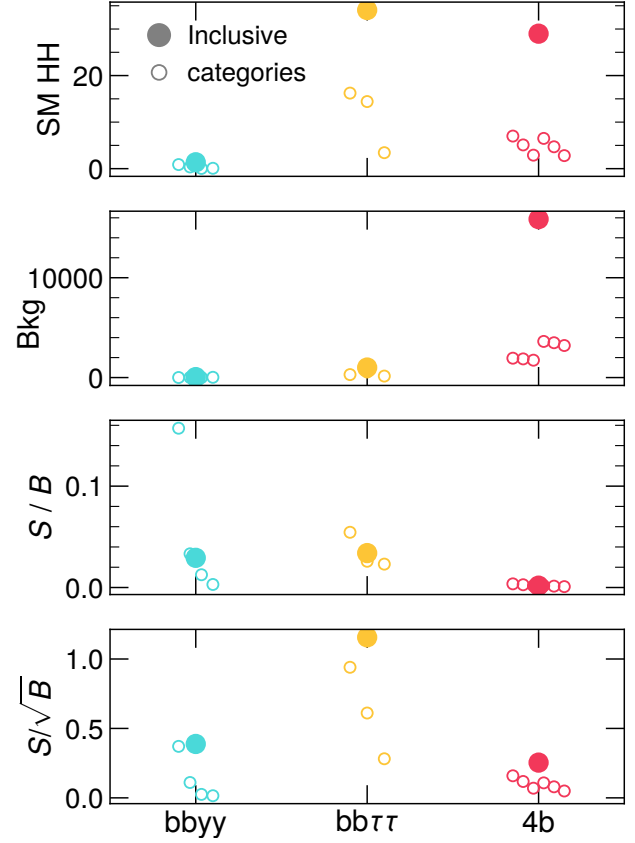


Figure C.8: Comparison of $bb\gamma\gamma$, $bb\tau\tau$, and $4b$ channels

	Observed	-2σ	-1σ	Expected	$+1\sigma$	$+2\sigma$
ggF production: ggF channel fit	5.73	4.44	5.97	8.28	12.51	19.71
ggF production: ggF+VBF channels fit	5.51	4.41	5.92	8.22	12.43	19.62
VBF production: VBF channel fit	122.9	72.0	96.7	134.3	194.2	281.6
VBF production: ggF+VBF channels fit	132.3	71.3	95.7	132.8	191.9	277.7
ggF+VBF production: VBF+ggF channels fit	5.45	4.34	5.83	8.09	12.21	19.19

Table C.8: The observed and expected upper limit on the SM HH production cross-section at the 95% CL. The expected value is shown with corresponding one and two standard deviation error bounds [7].

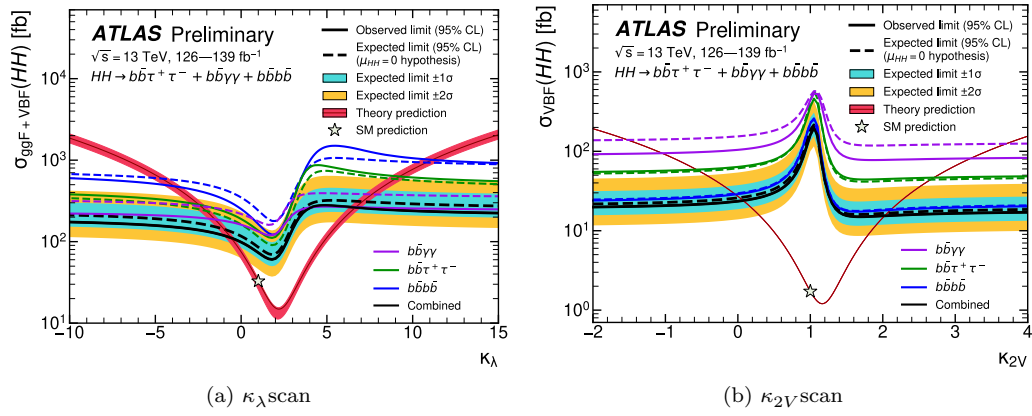


Figure C.7: Upper limits for the $b\bar{b}b\bar{b}$, $b\bar{b}\gamma\gamma$, and $b\bar{b}\tau^+\tau^-$ analysis channels and their combination [163].

Appendix D

3b1l category

D.0.1 Overview

For this iteration of the analysis, the possibility of including another ggF analysis category with 3b events was explored. To choose the last jet, the last jet passing the looser 85% b -tag efficiency working point, subsequently referred to as a “loose” b -tag (as motivated by studies in [184]). Including these events as another channel was explored to maximize the signal acceptance without compromising the sensitivity from increasing background yields. This new channel is called the “3b1l” category to refer to the 3b + 1 loose b -tag reconstruction.

Below is the suite of studies done for the 3b1l category and its subsequent impact when combining it with the 4b ggF analysis.

D.0.2 Event selection

The jet selection accuracy is 74% and 64% for the 4b and 3b1l ggF selection with %-level variations across the κ_λ values of interest. The 3b1l channel includes more noise due to looser b -tagging requirement, and therefore leads to a lower accuracy.

The corresponding $\Delta\eta_{HH}$ and X_{Wt} cuts distribution for this category is shown in Figure D.2.

The same SR as used as defined by the 4b ggF and VBF channels $X_{HH} \geq 1.6$ cut. Figure D.3(a) shows the impact of isolating the SM signal for the correctly paired events, and the corresponding Figure D.3(b) shows the isolation of this signal in the corresponding massplane. The corresponding massplane for the data events is shown in Figure D.3(c)

We have approximately the same amount of background events in the 3b1l channel as the ggF channel, but only about a fifth of the signal yield in 3b1l compared to 4b.

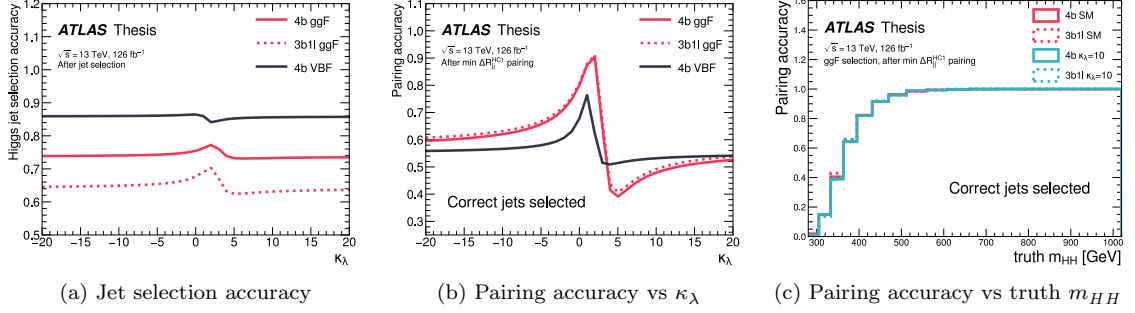
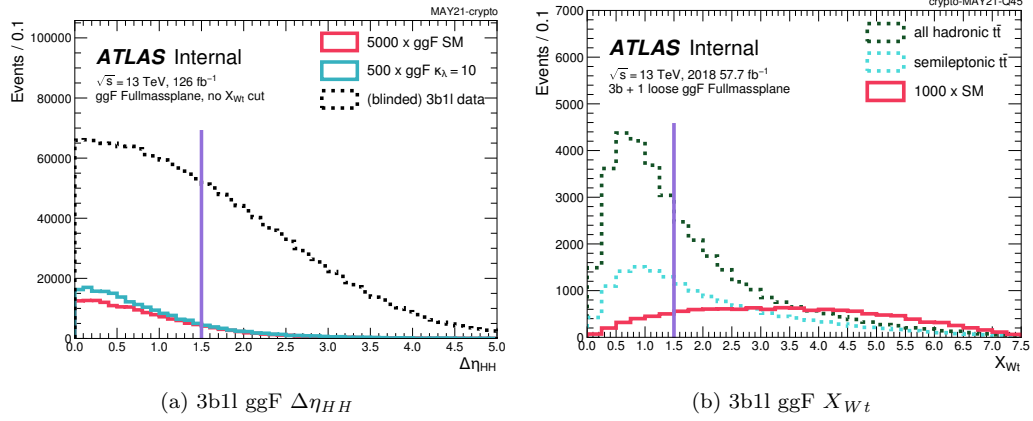


Figure D.1: The jet selection accuracy with the 3b1l category included.

Figure D.2: Visualization of the $\Delta\eta_{HH}$ and X_{Wt} cuts for the 3b1l channel with the ggF selection.

D.0.3 Background estimation

For the studies with the 3b1l category included, we split the 2b dataset into two subsets to use half the 2b stats for the 4b background estimate training and half the 2b stats for the 3b1l background estimate training. The corresponding validation plots are shown in Figure D.4 of the m_{HH} distribution in each year, broken down by the $\Delta\eta_{HH}$ cats. But overall, we saw a similarly good level of background modelling in 3b1l compared to 4b.

Due to loosened b -tagging requirement, we do have a much higher estimated $t\bar{t}$ contribution in our background, as demonstrated in Table D.1(a).

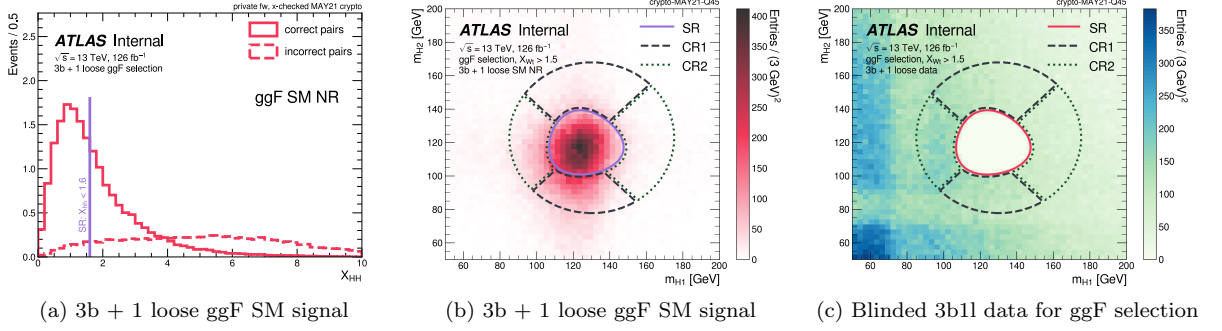


Figure D.3: 3b1l Signal Region definition showing the SM signal and (blinded) background distribution.

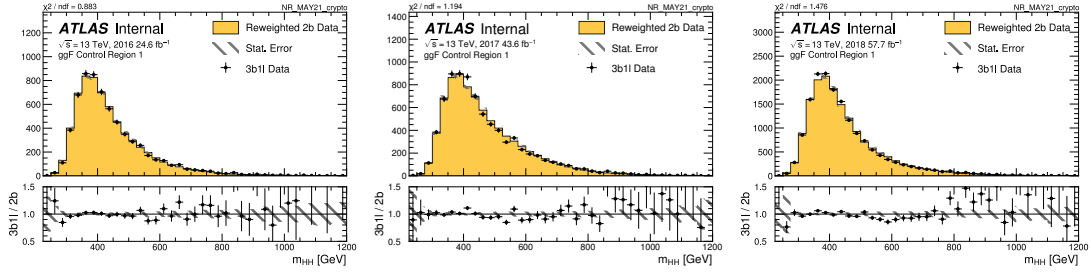


Figure D.4: CR1 after reweighting for the ggF 3b1l.

	16	17	18	All
pre- X_{Wt}	41.2%	39.6%	37.3%	38.7%
post- X_{Wt}	24.8%	23.5%	22.9%	23.5%

(a) ggF 3b1l

Table D.1: Percentage of the data-driven background estimate expected to be composed of $t\bar{t}$ events for the ggF 4b (left) and 3b1l categories (right).

D.0.4 Pre-fit plots and estimated impact on the limits

The same categorization used for the 4b ggF category was used for the 3b + 1 loose categories, and the visualization of the 3b + 1 loose histograms for each of the years is shown in Figure D.6. Similar qualitative trends as seen in the 4b histograms are seen in 3b + 1 loose, however in the significance subpanels, the significances are $\sim 5\times$ lower than for the 4b category due to the lower signal yield for the 3b + 1 loose events.

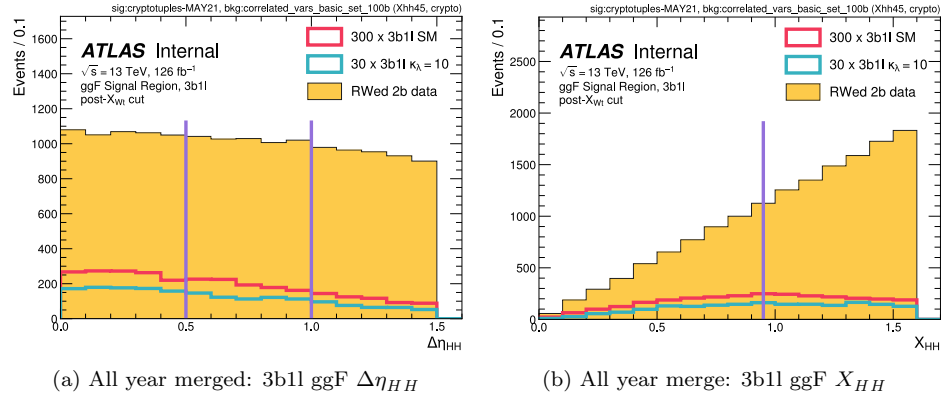
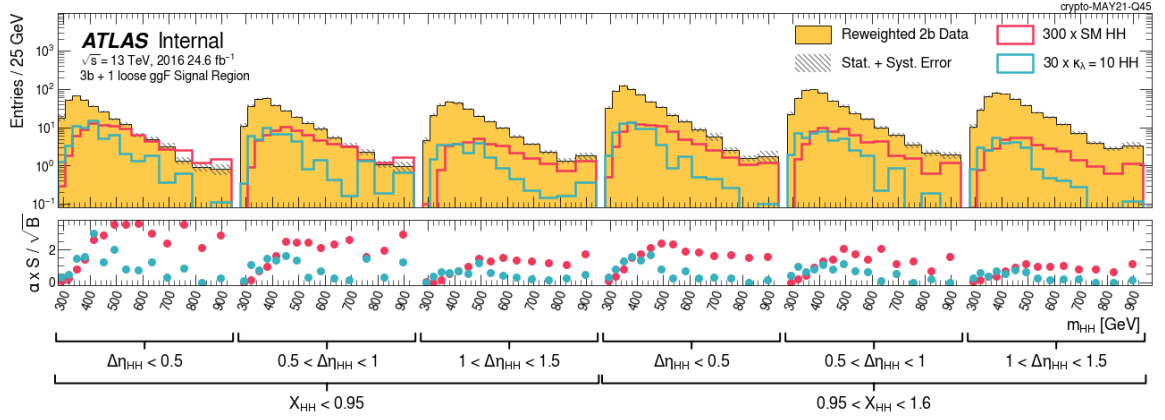
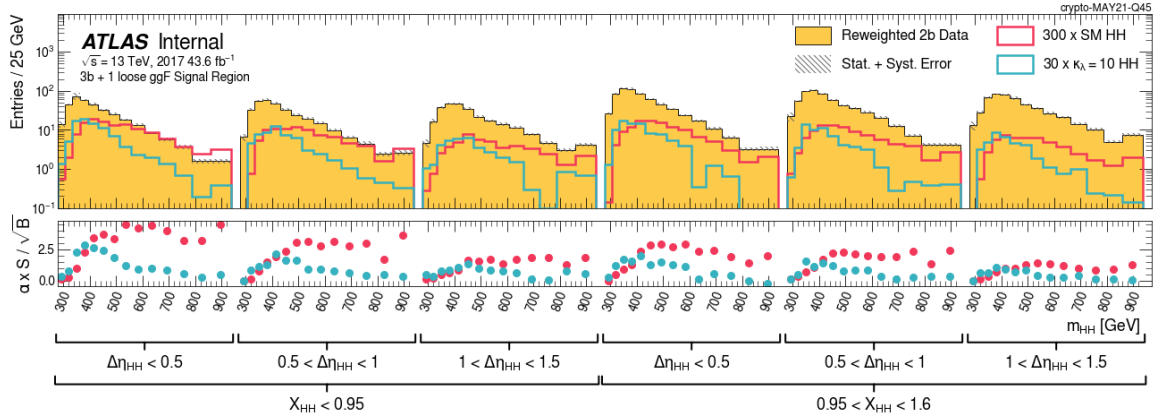


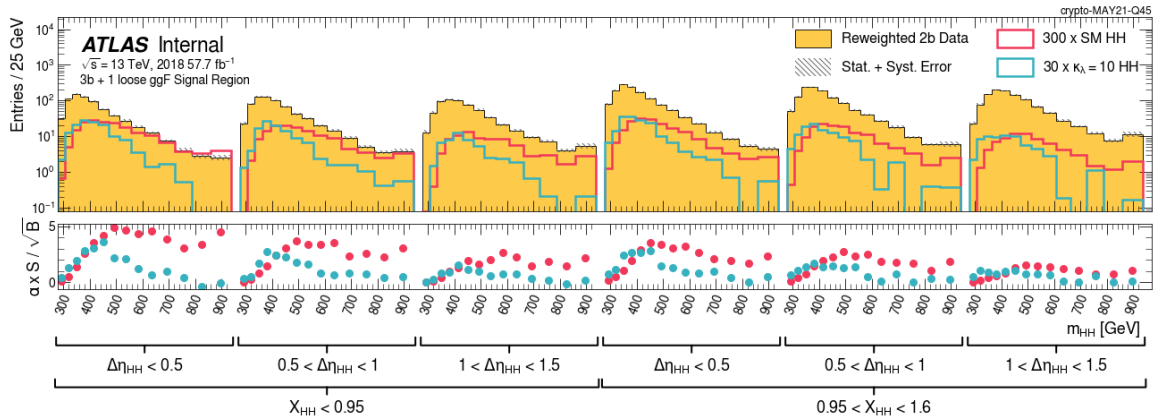
Figure D.5: Distributions of the variables used for categorization in 4b and 3b1l ggF. Years are merged. To visualize the signals they are scaled by $\alpha = 300$ and 30 for the SM NR and $\kappa_\lambda = 10$ signals, respectively.



(a) 2016: 3b + 1 loose ggF



(b) 2017: 3b + 1 loose ggF



(c) 2018: 3b + 1 loose ggF

Figure D.6: 3b + 1 loose ggF background and selected signal histograms for 2016, 2017, and 2018 with the proposed binning and categorization. To visualize the signals they are scaled by $\alpha = 300$ and 30 for the SM NR and $\kappa_\lambda = 10$ signals, respectively.

We then propagated the combined analysis selection with the 4b and 3b1l channels into a combined fit, and the results with uncorrelated nuisance parameters across the years, b-tag regions and the X_{HH} and $\Delta\eta_{HH}$ categories are shown in Figure D.7(a) for the κ_λ scan and the impact just on the SM limit highlighted in Figure D.7(b).¹ For the SM limit, by including the additional 3b + 1 loose selection a further $\sim 4\%$ is gained. Numerical results are listed in Table D.2, and showed only a small improvement on the constraints compared to the $\Delta\eta_{HH}$ and X_{HH} categorization. Therefore, our team opted not to include the 3b1l category to simplify the analysis. This allows the 4b ggF channel to use the full 2b statistics for a more robust background estimate, and correspondingly decrease the 2b stat error on our signal region prediction.

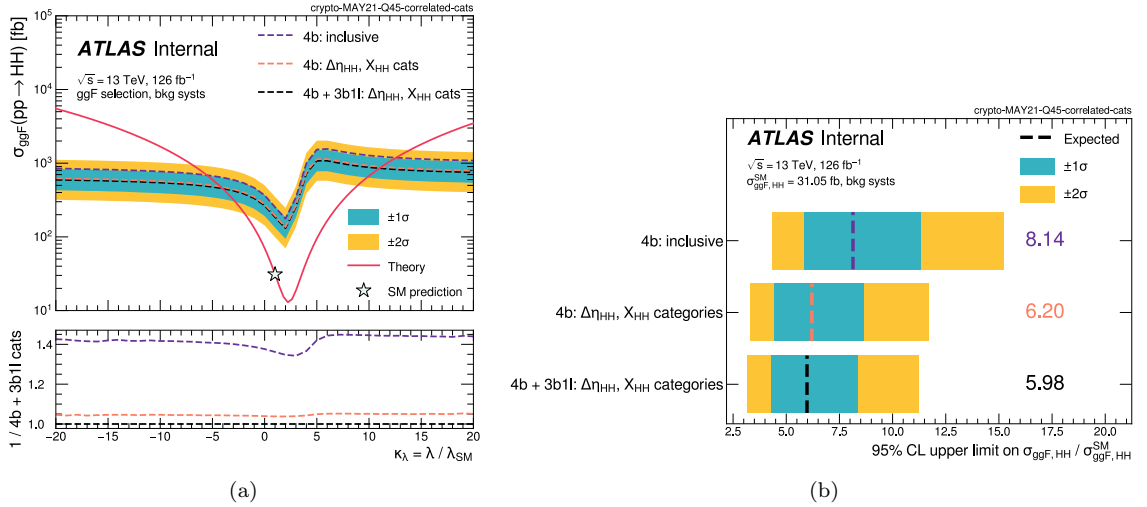


Figure D.7: The κ_λ scan with the ggF selection including the 4b and 3b + 1 loose selections and the $\Delta\eta_{HH}$ and X_{HH} categorization (left). Impact of the extra 3b1l category compared to just the ggF categorization (right). The background shape NPs are treated as uncorrelated across the X_{HH} and $\Delta\eta_{HH}$ categories.

Setup	κ_λ constraint	
4b: inclusive	[-5.52, 12.59]	
	correlated background NPs	uncorrelated background NPs
4b: $\Delta\eta_{HH}$, X_{HH} categories	[-4.21, 11.20]	[-3.98, 10.88]
4b + 3b1l: $\Delta\eta_{HH}$, X_{HH} categories	[-4.04, 11.00]	[-3.81, 10.70]

Table D.2: Impact on the κ_λ constraints for the various ggF categorization strategies.

¹ As before with the 4b limits, a finer m_{HH} binning than for the m_{HH} inclusive fit is used to maintain similar amount of bins with and without categorization.

Appendix E

Loss function derivations

Here we show the proofs of the functions that minimize the losses presented in Chapter 11.1. These proofs consider the infinite statistics limit and use the **Calculus of Variations**. A loss function in the form:

$$\mathcal{L}[f] = \int \mathcal{I}(x_1, \dots, x_n, f, f') dx. \quad (\text{E.1})$$

has extrema given by the Euler-Lagrange equation:

$$\frac{\partial \mathcal{I}}{\partial f} - \frac{d}{dx_i} \frac{\partial \mathcal{I}}{\partial f'} = 0$$

When there is no explicit dependence on the x_i , we just need to solve for $\frac{\partial \mathcal{I}}{\partial f} = 0$.

E.1 Cross-entropy loss function

The proof outlined below follows the one in [155]. The *logistic loss* for a classifier $D(x) \in (0, 1)$ is:

$$\mathcal{L}[D] = -\mathbb{E}_{x \sim p_{4b}} [\log D(x)] - \mathbb{E}_{x \sim p_{4b}} [\log (1 - D(x))] \quad (\text{E.2})$$

$$= \int [-\log D(x)p_{4b}(x) - \log(1 - D(x))p_{2b}(x)] dx \quad (\text{E.3})$$

We can use the calculus of variations to solve for the extrema of this integral. Let

$$\mathcal{I}(x_1, \dots, x_n, D, D') = -\log D(x)p_{4b}(x) - \log(1 - D(x))p_{2b}(x)$$

Since there is no explicit dependence on the x_i , solve for $\frac{\partial \mathcal{I}}{\partial w} = 0$.

$$\frac{\partial \mathcal{I}}{\partial D} = -\frac{p_{4b}}{D} + \frac{p_{2b}}{1-D} = 0 \implies \frac{p_{4b}}{p_{2b}} = \frac{D}{1-D} \implies \boxed{D = \frac{1}{1 - p_{4b}/p_{2b}}}$$

To show that this extremum is a minimum, we consider the second derivative.

$$\begin{aligned} \frac{\partial^2 \mathcal{I}}{\partial D^2} &= \frac{p_{4b}}{D^2} + \frac{p_{2b}}{(1-D)^2} \Big|_{D=\frac{1}{1-p_{4b}/p_{2b}}} \\ &= p_{4b} \left(1 + \frac{p_{4b}}{p_{2b}}\right) + p_{2b} \left(\frac{1 + p_{4b}/p_{2b}}{p_{4b}/p_{2b}}\right) \\ &= p_{4b} \left(1 + \frac{p_{4b}}{p_{2b}}\right) + p_{2b} \left(1 + \frac{p_{2b}}{p_{4b}}\right) > 0, \end{aligned}$$

as both the probabilities p_{4b} and p_{2b} are by definition non-negative. The loss is concave up at $D = 1/(1 - p_{4b}/p_{2b})$, and this D minimizes the loss.

E.2 Reweighting loss function

The derivation below follows the proof from [185].

In Chapter 11.1 we introduced the loss function:

$$\mathcal{L}[Q] = \mathbb{E}_{x \sim p_{2b}} \left[\exp \left(\frac{1}{2} Q(x) \right) \right] + \mathbb{E}_{x \sim p_{4b}} \left[\exp \left(-\frac{1}{2} Q(x) \right) \right] \quad (\text{E.4})$$

We want to show that $Q^*(x) = \log w^*(x)$ minimizes this loss for $w^*(x) = p_{4b}(x)/p_{2b}(x)$. Since \log is a (monotonically increasing) bijective function, we can just show that the loss function with respect to w is minimized when $w^*(x) = p_{4b}(x)/p_{2b}(x)$. Substituting $Q(x) = \log w(x)$ into the loss function and writing out the integrals defining these expectation values, we have:

$$\mathcal{L}[w] = \mathbb{E}_{x \sim p_{2b}} \left[\sqrt{w(x)} \right] + \mathbb{E}_{x \sim p_{4b}} \left[\frac{1}{\sqrt{w(x)}} \right] \quad (\text{E.5})$$

$$= \int \left[\sqrt{w(x)} p_{2b}(x) + \frac{1}{\sqrt{w(x)}} p_{4b}(x) \right] dx. \quad (\text{E.6})$$

We again use the calculus of variations to solve for the extrema of this integral. Let

$$\mathcal{I}(x_1, \dots, x_n, w, w') = \sqrt{w(x)} p_{2b}(x) + \frac{1}{\sqrt{w(x)}} p_{4b}(x).$$

Solve for $\frac{\partial \mathcal{I}}{\partial w} = 0$.

$$\begin{aligned}\frac{\partial \mathcal{I}}{\partial w} &= \frac{1}{2}w^{-1/2}p_{2b} - \frac{1}{2}w^{-3/2}p_{4b} = 0 \\ \implies w^*(x) &= p_{4b}/p_{2b}\end{aligned}$$

This is the w^* that we were searching for! To show that it's also a minimum, we take the 2^{nd} functional derivative:

$$\frac{\partial^2 \mathcal{I}}{\partial w^2} = -\frac{1}{4}w^{-3/2}p_{2b} + \frac{3}{4}w^{-5/2}p_{4b} = \frac{1}{4}w^{-5/2} \left(3p_{4b} - \frac{1}{4}wp_{2b} \right)$$

and evaluate it at the extrema point:

$$\left. \frac{\partial^2 \mathcal{I}}{\partial w^2} \right|_{w=p_{4b}/p_{2b}} = \frac{1}{4} \left(\frac{p_{4b}}{p_{2b}} \right)^{-5/2} \left(\frac{11}{4}p_{4b} \right) > 0$$

because both $p_{4b}(x)$ and $p_{2b}(x)$ are positive. $\mathcal{L}[w]$ is concave up at $w^*(x) = p_{4b}(x)/p_{2b}(x)$ and this w^* is in fact a minimum. This justifies our use of stochastic gradient descent with this loss function to derive these reweighting maps.

Appendix F

$t\bar{t}$ aware reweighting

F.1 Motivation

As shown in Table 11.2, $t\bar{t}$ accounts for 10% of the $4b$ background. The background estimate inclusively reweights both the QCD and $t\bar{t}$ contributions, but we can use the $t\bar{t}$ simulation to evaluate the data trained reweighting on the $2b$ $t\bar{t}$ simulation and compare the performance to the $4b$ $t\bar{t}$. Figure F.1 shows this result for the m_{HH} modeling, and the inclusively trained background estimate over-predicts all-had $t\bar{t}$ by **2.28** and semi-leptonic $t\bar{t}$ by **3.4** since the proportion of true $2b$ events is higher in $t\bar{t}$. The subpanels in Figure F.1 have a flat ratio for $4b$ $t\bar{t}$ over the reweighted $2b$ prediction, so the inclusive reweighting is accurately modeling the $t\bar{t}$ shape.

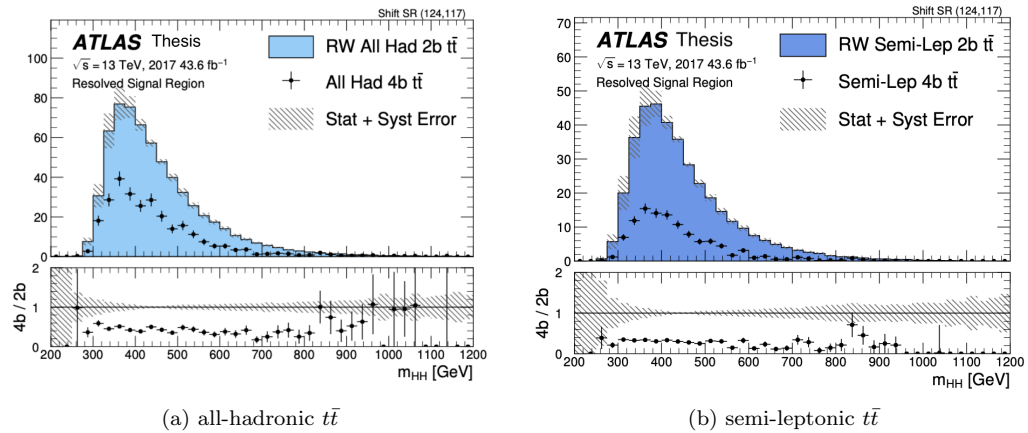


Figure F.1: Performance of the inclusively trained reweighting evaluated on the $t\bar{t}$ simulation. The performance of the inclusively trained reweighting is evaluated on $2b$ $t\bar{t}$ simulation and compared to the $4b$ $t\bar{t}$ prediction. The error bar on the background prediction shows the quadrature sum of the $2b$ Poisson, deep ensembles, and CR1 / CR2 shape systematic error.

This motivated us to construct a more accurate background model by treating the $t\bar{t}$ and QCD components separately. As emphasized in Figure 11.1 – QCD is the contribution where we don’t trust our simulation and need to resort to data-driven methods, and Figure F.2 shows how we plan to use the data-driven model just for the QCD contribution.

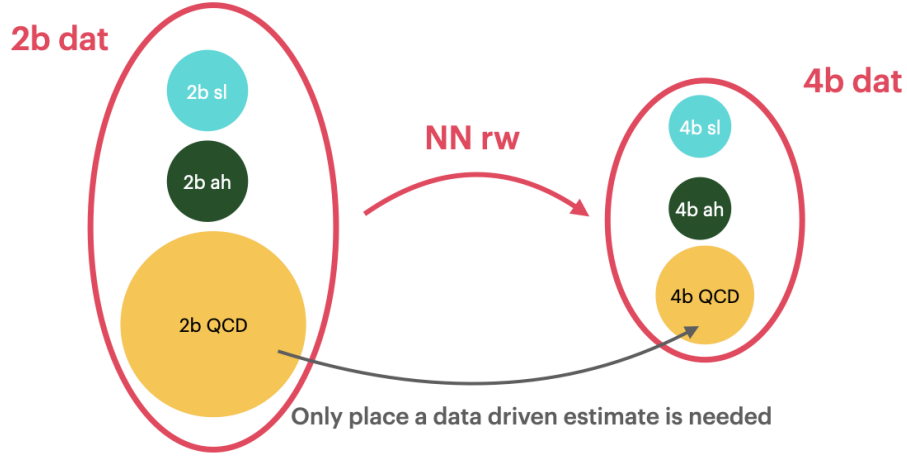


Figure F.2: Illustration of how the data-trained reweighting (in pink) inclusively reweights all of the background components. The only background component that we need to use a data-driven technique for is the QCD piece where we don’t trust our simulation.

Modeling the components separately can also decrease the total normalization error. This is because for two random variables, X and Y , if $Z = X + Y$, then $\text{Cov}[Z] = \text{Cov}[X] + \text{Cov}[Y] + 2\text{Cov}[X, Y]$, i.e., a negative $\text{Cov}[X, Y]$ decreases the error on Z . If the $t\bar{t}$ and QCD normalizations are negatively correlated, this will improve the background model by decreasing the predicted error.

This appendix has two parts. In Appendix F.2 studies of fitting the $t\bar{t}$ and QCD templates with the data trained estimate as the QCD template are presented, while in Appendix F.3 preliminary studies are shown for a novel method deriving a pure QCD reweighting – which is then used for the QCD template.

F.2 Fitting $t\bar{t}$ templates

F.2.1 Prescription

The 36 fb^{-1} analysis was able to model QCD separately through the iterative 1d corrections. They then fit the QCD and $t\bar{t}$ template normalizations in dedicated top and QCD-enriched CRs, and the fitting procedure described here follows closely what in [116]. Since X_{Wt} is an important discriminating variable, for these studies the background estimate is trained *before* applying the $X_{Wt} > 1.5$ cut.

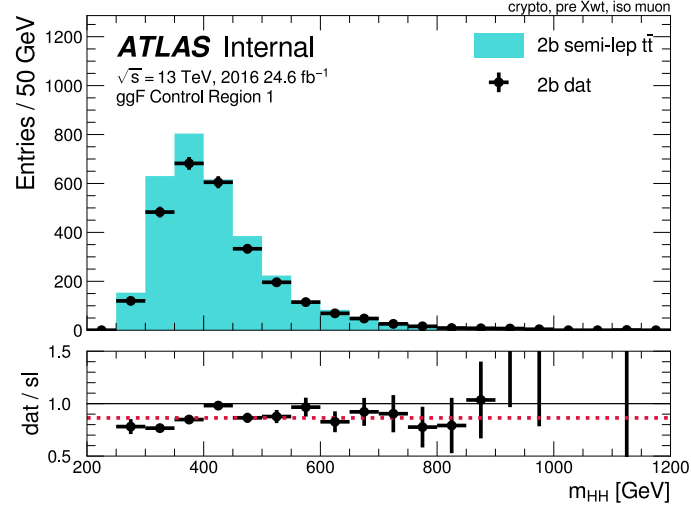


Figure F.3: m_{HH} in $2b$ sample with an isolated muon for the 2016 data in CR1. The dashed red line in the subpanel shows the fitted $\alpha_{t\bar{t}, 2b}^{sl}$

The QCD template is defined from the reweighted $2b$ data with the contributions from the reweighted $2b$ $t\bar{t}$ samples subtracted off:

$$\text{QCD} = \text{NN}(2b \text{ dat}) - \text{NN}(2b \text{ all-had } t\bar{t}) - \alpha_{t\bar{t}, 2b}^{sl} \cdot \text{NN}(2b \text{ semi-lep } t\bar{t}) \quad (\text{F.1})$$

To correct the normalization for the semi-leptonic $t\bar{t}$ contribution in the sub, we additionally require an isolated muon to define a SF:

$$\alpha_{t\bar{t}, 2b}^{sl} = \frac{\text{2b semi-leptonic } t\bar{t} \text{ with iso } \mu}{2b \text{ data with iso } \mu}$$

An illustration of this fit in the 2016 CR1 is shown in Figure F.3. The m_{HH} histogram for the data and semi-leptonic $t\bar{t}$ histograms are both shown, but their ratio is nearly flat, and closely follows the red dashed line of the inclusive $\alpha_{t\bar{t}, 2b}^{sl}$ fit. A comparison of the $\alpha_{t\bar{t}, 2b}^{sl}$ for the rest of the years and also in CR2 is shown in Table F.2.

$\alpha_{t\bar{t}, 2b}^{sl}$	2016	2017	2018
CR1	0.86	0.80	0.83
CR2	0.82	0.81	0.79

Table F.1: Fitted SFs for $\alpha_{t\bar{t}, 2b}^{sl}$.

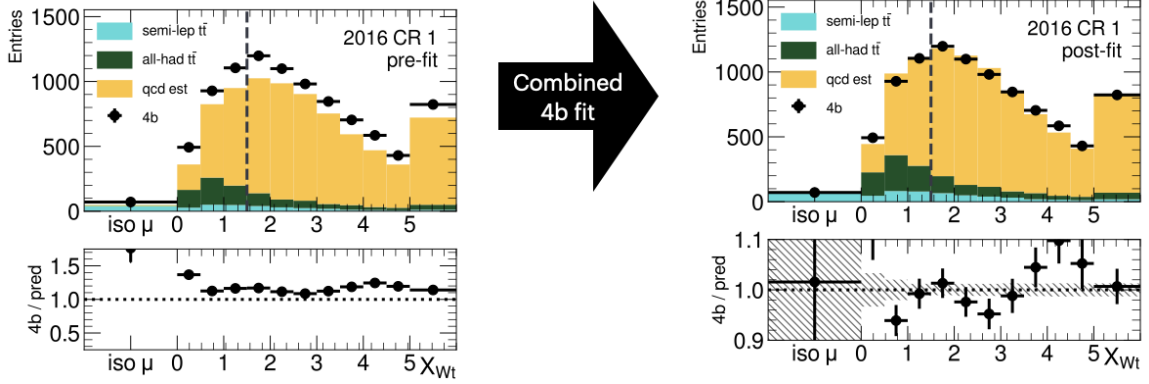


Figure F.4: The first bin shows 4b events with an isolated muon and the rest of the distribution shows the X_{Wt} histogram. Overflow events are included in the highest bin. The left distribution shows the pre-fit templates and the right distribution shows the post-fit templates with optimized normalization. The statistical errors of the fit are shown in the hatched grey lines on the right subpanel. $X_{Wt} = 1.5$ (analysis cut) is also shown in the figures.

With the QCD template from Eq. F.1 and the $t\bar{t}$ templates from the 4b MC, the normalizations for these templates are determined by a combined 4b fit with:

- the X_{Wt} shape (helps discriminate between the all-hadronic $t\bar{t}$ and qcd templates), and
- the 4b data with an isolated μ (again constrains the semi-leptonic $t\bar{t}$ component).

This discriminant in the 4b CR1 is shown in Figure F.4, with left figure showing the pre-fit templates and the right figure showing the post-fit templates with good agreement with the 4b data.

F.2.2 Fit results 4b

Figure F.5 shows the correlation between the fitted normalizations for the combined 4b fit in CR1 (top) and CR2 (bottom). The QCD and all had templates have a strong negative correlation (varying between -0.62 and -0.65) indicating that including these as separate contributions in the fit should decrease the overall normalization. The semi-leptonic $t\bar{t}$ normalization has a smaller correlation with the other samples: -0.3 with respect to the all hadronic $t\bar{t}$ and -0.1 with respect to the QCD.

Figure F.6 aggregates the information across the fits in both CR1 and CR2. The first row “CR1 SF” shows the fitted normalizations (along with the statistical uncertainty of the fit) in CR1, corresponding to the fit in Figure F.5. The second row “CR2” shows the analogous result, but for the fits conducted in CR2. The third row “extrap norm” is the ratio of the SFs in CR2 and CR1 ... and the uncertainty that we assess on the QCD template this was less than a % – almost an order of magnitude smaller than the shape uncertainty from Table 11.3. This is also compared to the ratio between the CR1 and CR2 background predictions in the SR (last row) – and the difference from unity is quite a bit larger than the difference of the SFs from the QCD templates.

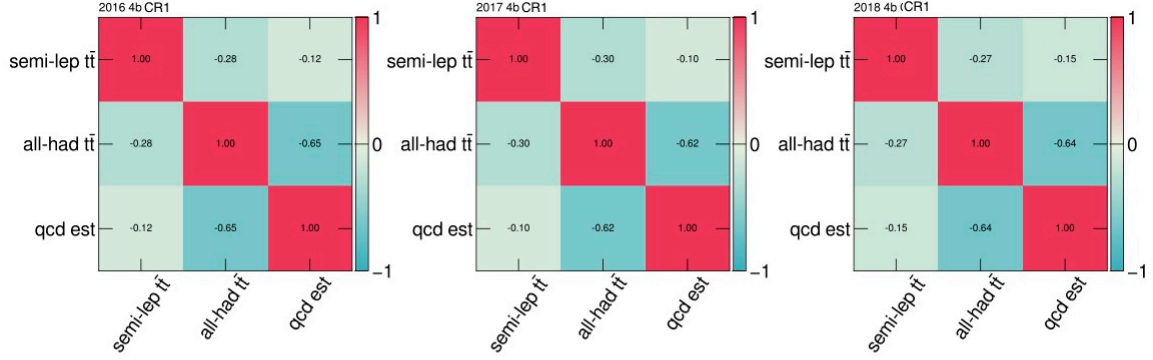


Figure F.5: The correlation between the fitted template normalizations from the combined 4b fit in CR1. The columns correspond to the different years of data taking.

	2016			2017			2018		
	semi-lep	all-had	QCD	semi-lep	all-had	QCD	semi-lep	all-had	QCD
CR 1 SF	1.640 ± 0.176	1.322 ± 0.131	1.113 ± 0.020	1.362 ± 0.139	1.740 ± 0.114	1.096 ± 0.016	1.075 ± 0.115	1.673 ± 0.088	1.082 ± 0.013
CR 2 SF	1.841 ± 0.200	1.456 ± 0.135	1.106 ± 0.020	1.540 ± 0.141	1.806 ± 0.110	1.103 ± 0.016	1.232 ± 0.110	1.665 ± 0.083	1.082 ± 0.013
extrap norm CR 2 SF / CR1 SF	1.122	1.101	0.9933	1.130	1.0379	1.0062	1.147	0.995	1.00044
naive extrap norm CR 2 pred / CR1 pred	0.992			0.996			0.968		

Figure F.6: Evaluating an uncertainty based on the 4b normalization fits in CR1 and CR2.

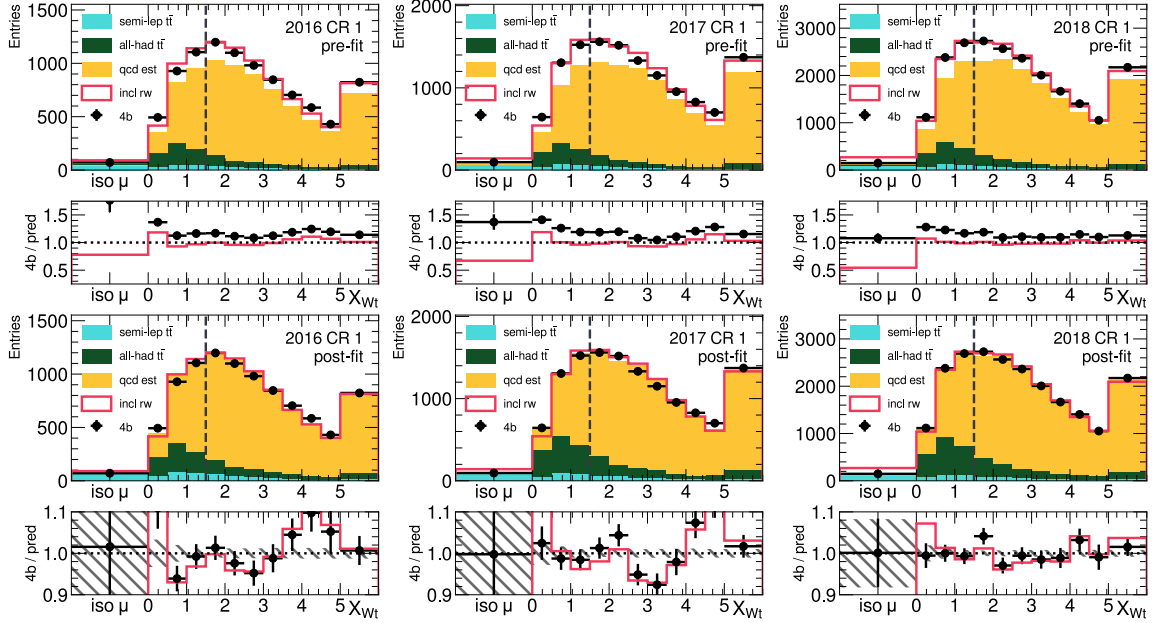


Figure F.7: CR1 fit in CR1. The top rows show the pre-fit templates and the bottom rows show the post-fit norms for the templates. The statistical errors of the fit are shown in the hatched grey lines on the right sub-panel. The pink line shows the inclusively trained and evaluated reweighting (the baseline analysis strategy).

Despite these promising normalization constraints, when the post-fit X_{Wt} plot was compared to the inclusive reweighting (without these component fits), these approaches had nearly identical results for the $X_{Wt} > 1.5$ region, as shown in the bottom row of Figure F.7. This means that (for the current dataset size) we’re not seeing a meaningful difference from this approach for fitting the components separately.

A comparison of the m_{HH} modeling for all of the years in CR1 after the $X_{Wt} > 1.5$ cut is shown in Figure F.8. The first subpanel compares the non-closure of the two predictions (template fit in black and the inclusive reweighting prediction in pink) – and we see nearly identical performance up to 800 GeV – the bulk of the distribution that drives the analysis’s performance. The bottom sub-panel just compares these two predictions. The template fit prediction has $\mathcal{O}(5\%)$ fewer events in the turn-on region (below 300 GeV, and also predicts more events in the high m_{HH} tail. Overall the predictions are quite similar.

Figure F.9 shows the m_{HH} distribution with the same CR1 normalizations, but evaluated in the CR2 control region, which allows a test of the normalization fits extrapolation. We see the same messages as in the CR1 evaluation in Figure F.8, and although it was disappointing that new method did not do better here – this is a nice check of the robustness of the normalization fits in the extrapolation.

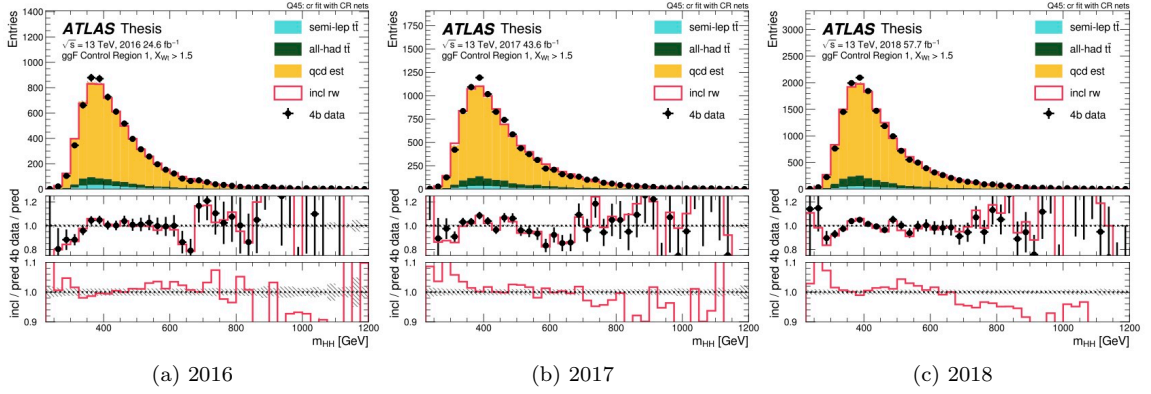


Figure F.8: CR1 evaluation using the template normalizations from the CR1 fits with the $X_{Wt} > 1.5$ cut. The first subpanel shows the ratio of the 4b to the background prediction, with the black dots showing the ratio to the template fits, and the pink curve showing the ratio to the inclusive reweighting prediction. The bottom panel compares the ratio of the inclusive reweighting to the template fit prediction.

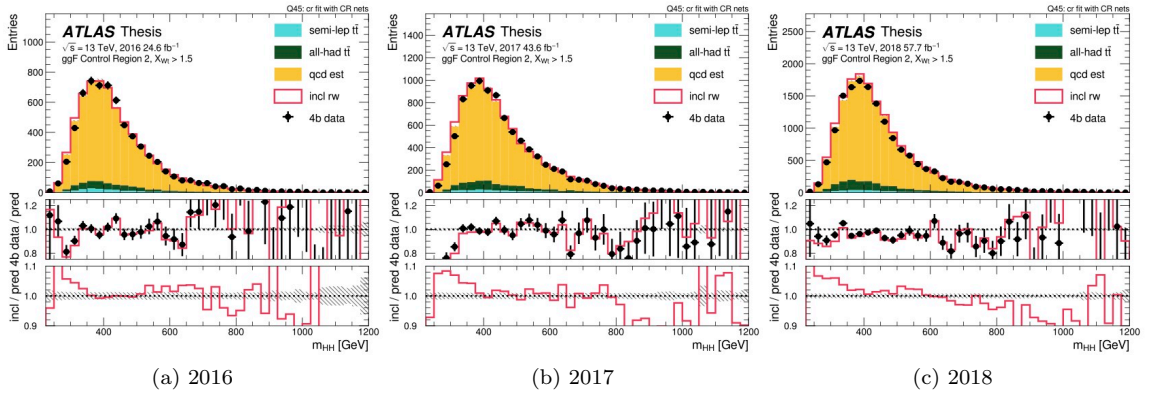


Figure F.9: CR2 evaluation using the template normalizations from the CR1 fits. The first subpanel shows the ratio of the 4b to the background prediction, with the black dots showing the ratio to the template fits, and the pink curve showing the ratio to the inclusive reweighting prediction. The bottom panel compares the ratio of the inclusive reweighting to the template fit prediction.

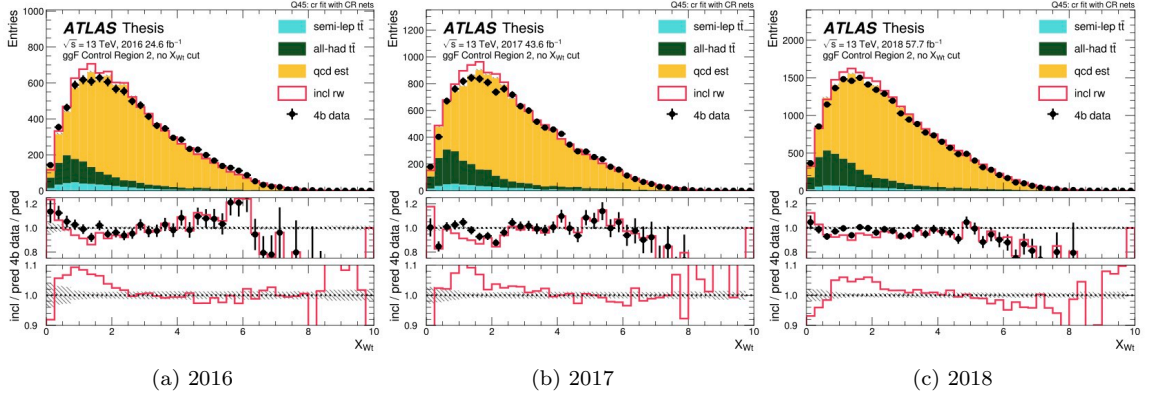


Figure F.10: CR2 evaluation using the template normalizations from the CR1 fits with the $X_{Wt} > 1.5$ cut. The first subpanel shows the ratio of the 4b to the background prediction, with the black dots showing the ratio to the template fits, and the pink curve showing the ratio to the inclusive reweighting prediction. The bottom panel compares the ratio of the inclusive reweighting to the template fit prediction.

Figure F.10 shows the X_{Wt} CR2 distributions, again with the CR1 template normalizations. Similar to what was seen in CR1, the X_{Wt} closure is improved by 5 – 10% in the $X_{Wt} < 1.5$ region.

F.2.3 Fit results 3b1f

These tests were also repeated in the 3b1f region and the extrapolation into the SR was checked. Figure F.11 shows the fitted template normalizations in both CR1 and CR2, and we again see a sub-percent level constraint on the QCD template normalization, which is also reflected in the m_{HH} $X_{Wt} > 1.5$ SR in Figure F.13. Figure F.14 shows the X_{Wt} shape, and although the closure is slightly better in the lowest X_{Wt} bins – the results are largely compatible.

Figure F.12 shows the result of applying these fitted CR1 normalizations in on the CR1 discriminants. Similar to the 4b result, the template fit prediction is largely compatible with the inclusive fit prediction for $X_{Wt} > 1.5$.

	2016			2017			2018		
	semi-lep	all-had	QCD	semi-lep	all-had	QCD	semi-lep	all-had	QCD
CR SF	0.995 ± 0.036	1.000 ± 0.033	1.003 ± 0.006	0.916 ± 0.033	1.740 ± 0.114	1.002 ± 0.006	0.929 ± 0.023	1.044 ± 0.022	0.993 ± 0.004
VR SF	0.955 ± 0.034	1.017 ± 0.030	1.004 ± 0.006	0.930 ± 0.032	1.806 ± 0.110	1.009 ± 0.006	0.906 ± 0.022	1.044 ± 0.020	0.994 ± 0.004
extrap norm	0.960	1.018	1.0013	1.015	0.971	1.0069	0.975	0.995	1.0014
extrap norm naive	0.970			0.963			0.971		

Figure F.11: Evaluating an uncertainty based on the 3b1f normalization fits in CR1 and CR2.

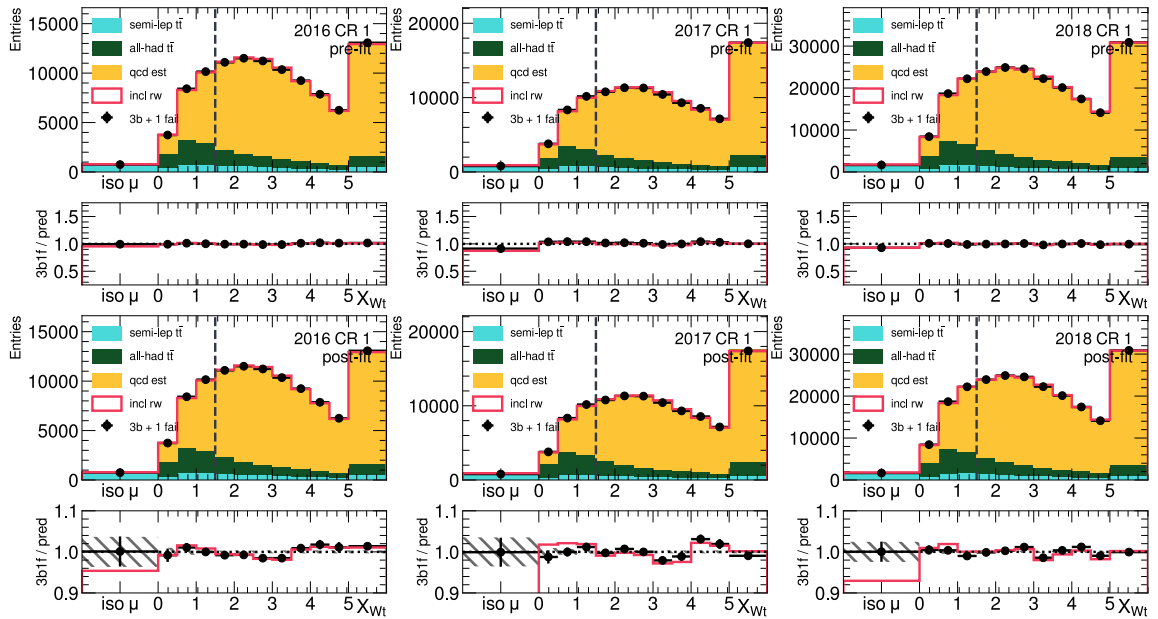


Figure F.12: CR1 fit in CR1

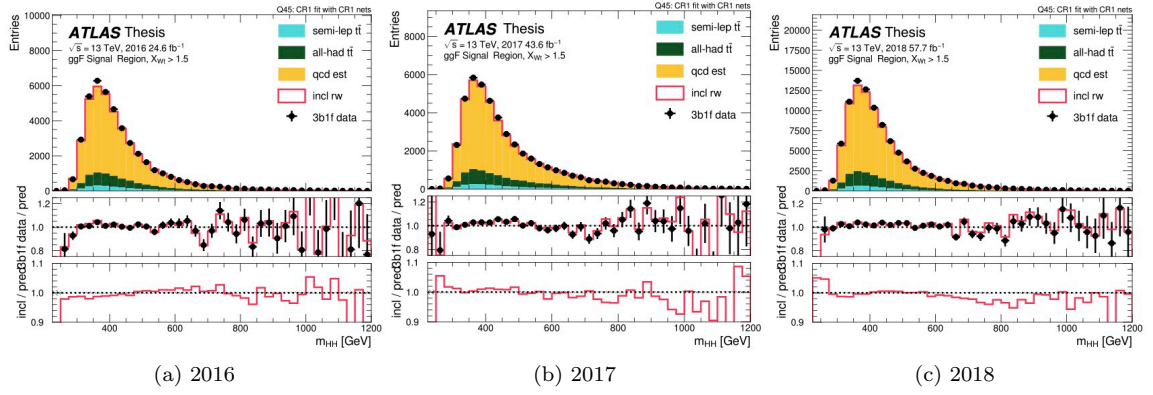


Figure F.13: 3b1f SR evaluation using the CR1 fits

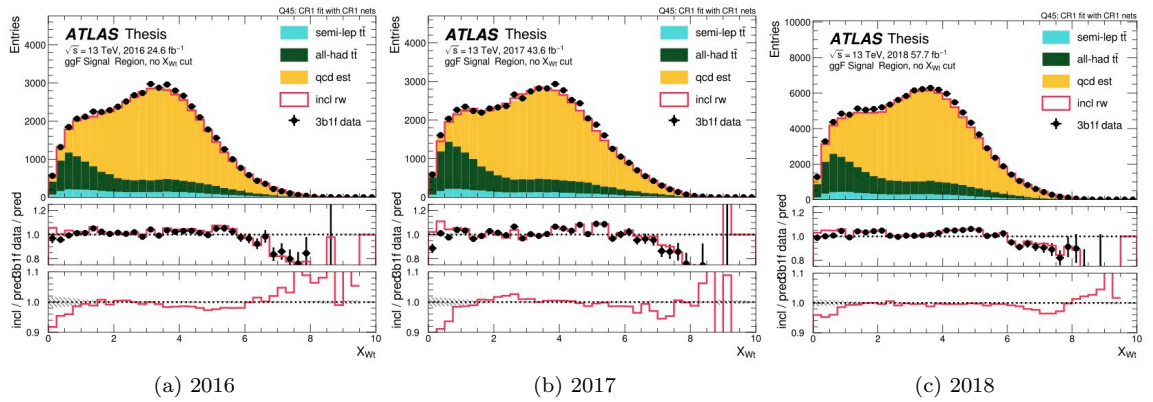


Figure F.14: 3b1f SR evaluation using the CR1 fits

F.3 Pure QCD reweighting

In the previous section, the template fits failed to outperform the inclusive reweighting. However, in those studies the QCD template was approximated by the inclusively trained reweighting. In this section, we show how to modify the reweighting definition to derive a pure QCD reweighting.

F.3.1 Mathematical formulation

Rafael Teixeira de Lima developed this formalism, and this write-up closely follows his notes [186].

Let $R_{2b^{all} \rightarrow 4b^{all}}$ describe the inclusively trained reweighting:

$$R_{2b^{all} \rightarrow 4b^{all}} = \frac{\alpha_{4b}^{t\bar{t}} p_{4b}^{t\bar{t}} + \alpha_{4b}^Q p_{4b}^Q}{\alpha_{2b}^{t\bar{t}} p_{2b}^{t\bar{t}} + \alpha_{2b}^Q p_{2b}^Q} \quad (\text{F.2})$$

where p denotes the probability densities and α the corresponding normalizations. The subscripts $2b$ and $4b$ indicate the b -tagging requirement and the superscripts indicate the $t\bar{t}$ and QCD (abbreviated “Q”) samples. We simplify the notation by letting $P_x^y = \alpha_x^y p_x^y$ so that Eq. F.2 becomes:

$$\begin{aligned} R_{2b^{all} \rightarrow 4b^{all}} &= \frac{P_{4b}^{t\bar{t}} + P_{4b}^Q}{P_{2b}^{t\bar{t}} + P_{2b}^Q} \\ &= \frac{P_{4b}^{t\bar{t}}}{P_{2b}^{t\bar{t}} + P_{2b}^Q} + \frac{P_{4b}^Q}{P_{2b}^{t\bar{t}} + P_{2b}^Q} \end{aligned}$$

We next want to isolate the individual reweightings so that we can solve for P_{4b}^Q/P_{2b}^Q :

$$\begin{aligned} R_{2b^{all} \rightarrow 4b^{all}} &= \frac{P_{4b}^{t\bar{t}}}{P_{2b}^{t\bar{t}}} \cdot \frac{1}{1 + \frac{P_{2b}^Q}{P_{2b}^{t\bar{t}}}} + \frac{P_{4b}^Q}{P_{2b}^Q} \cdot \frac{1}{1 + \frac{P_{2b}^{t\bar{t}}}{P_{2b}^Q}} \\ &= \frac{P_{4b}^{t\bar{t}}}{P_{2b}^{t\bar{t}}} \cdot \frac{P_{2b}^{t\bar{t}}}{P_{2b}^{t\bar{t}} + P_{2b}^Q} + \frac{P_{4b}^Q}{P_{2b}^Q} \cdot \frac{P_{2b}^Q}{P_{2b}^Q + P_{2b}^{t\bar{t}}} \\ &= \frac{P_{4b}^{t\bar{t}}}{P_{2b}^{t\bar{t}}} \cdot \frac{P_{2b}^{t\bar{t}}}{P_{2b}^{t\bar{t}} + P_{2b}^Q} + \frac{P_{4b}^Q}{P_{2b}^Q} \cdot \left(1 - \frac{P_{2b}^{t\bar{t}}}{P_{2b}^Q + P_{2b}^{t\bar{t}}}\right) \end{aligned} \quad (\text{F.3})$$

We can now identify:

- $R_{2b^{t\bar{t}} \rightarrow 4b^{t\bar{t}}} = \frac{P_{4b}^{t\bar{t}}}{P_{2b}^{t\bar{t}}}$: An MC based $2b \rightarrow 4b$ $t\bar{t}$ reweighting,
- $R_{2b^{t\bar{t}} \rightarrow 4b^{t\bar{t}}} = \frac{P_{2b}^{t\bar{t}}}{P_{2b}^Q + P_{2b}^{t\bar{t}}}$: reweighting the $2b$ data into the $2b$ $t\bar{t}$.
- $R_{2b^Q \rightarrow 4b^Q}$: the pure QCD reweighting – what we want to solve for.

The first two reweightings in the list above are ones that we can easily derive. $R_{2b^{all} \rightarrow 4b^{all}}$ is the current background estimate, so we can move forward to solve for $R_{2b^Q \rightarrow 4b^Q}$. First, substitute these reweighting definitions into Eq. F.3:

$$R_{2b^{all} \rightarrow 4b^{all}} = R_{2b^{t\bar{t}} \rightarrow 4b^{t\bar{t}}} \cdot R_{2b^{t\bar{t}} \rightarrow 4b^{t\bar{t}}} + R_{2b^Q \rightarrow 4b^Q} \left(1 - R_{2b^{t\bar{t}} \rightarrow 4b^{t\bar{t}}}\right), \quad (\text{F.4})$$

and then rearrange to solve for $R_{2b^Q \rightarrow 4b^Q}$:

$$R_{2b^Q \rightarrow 4b^Q} = \frac{R_{2b^{all} \rightarrow 4b^{all}} - R_{2b^{t\bar{t}} \rightarrow 4b^{t\bar{t}}} \cdot R_{2b^{t\bar{t}} \rightarrow 4b^{t\bar{t}}}}{1 - R_{2b^{t\bar{t}} \rightarrow 4b^{t\bar{t}}}}. \quad (\text{F.5})$$

Now that we have a pure QCD reweighting, this gives a prescription for modeling the 4b data:

$$P_{4b}^{all} = R_{2b^Q \rightarrow 4b^Q} \cdot \left(P_{2b}^{all} - P_{2b}^{t\bar{t}}\right) + P_{4b}^{t\bar{t}}. \quad (\text{F.6})$$

The first term $R_{2b^Q \rightarrow 4b^Q} \cdot \left(P_{2b}^{all} - P_{2b}^{t\bar{t}}\right)$ gives the the QCD template, where the NN weights from Eq. F.5 are used. To get just the QCD template, we evaluate this reweighting on the 2b data and then subtract off the result of applying the reweighting to the 2b $t\bar{t}$ MC. The second term $P_{4b}^{t\bar{t}}$ is the $t\bar{t}$ template – given by the $t\bar{t}$ MC.

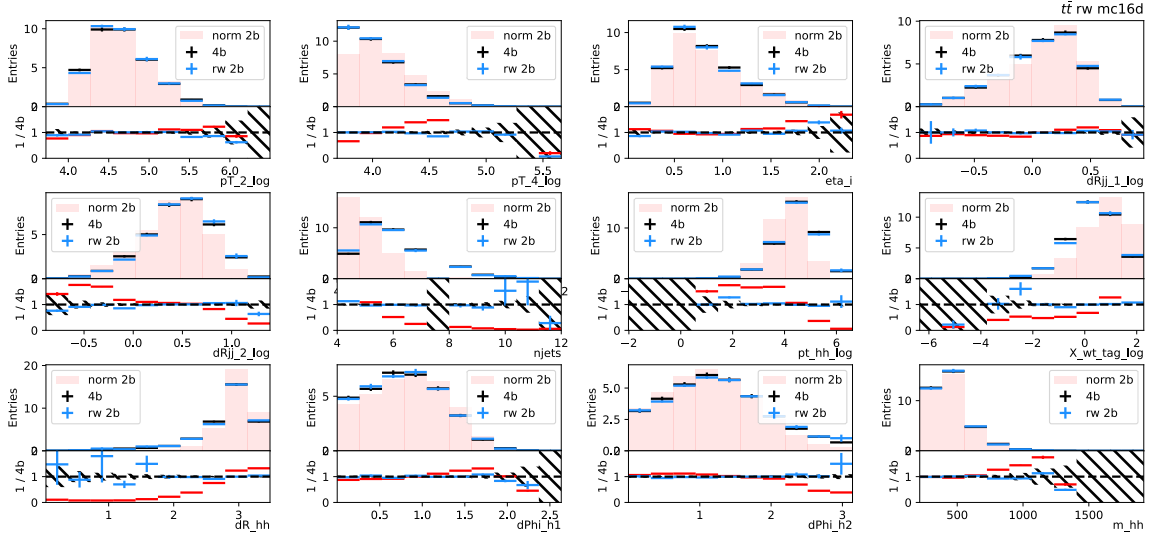
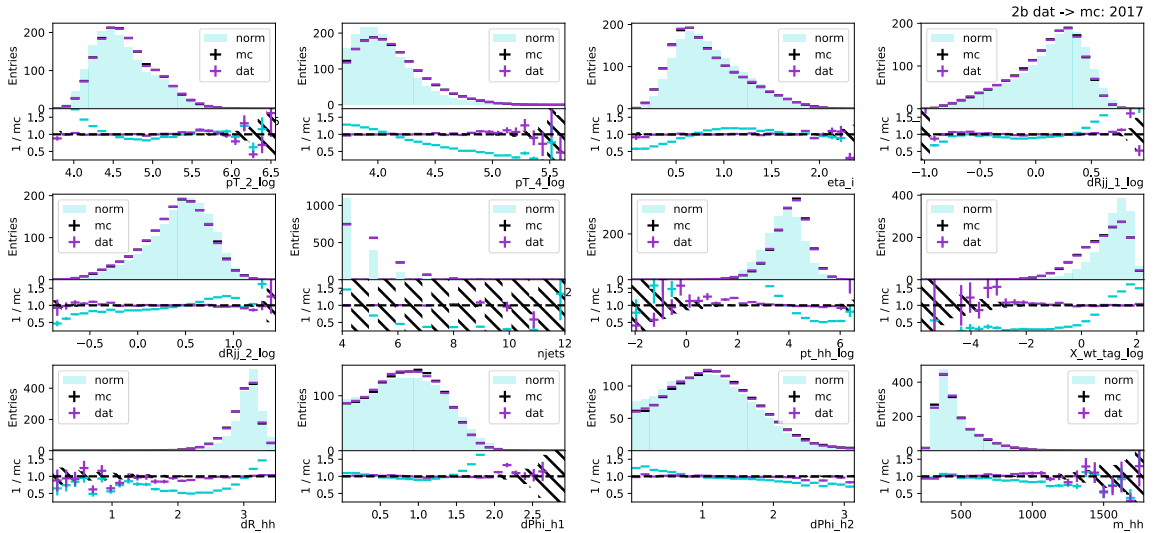
F.3.2 Experiments

The following reweighting estimates show a single training (instead of the machinery with 100 bootstraps) to get a first sense of how much this method could help us. We will fit the X_{W_t} template again, so the reweightings are derived in CR1 before the X_{W_t} cut.

Using the NN setup from the inclusive reweighting (3 hidden layers with 50 neurons each) for the $R_{2b^{t\bar{t}} \rightarrow 4b^{t\bar{t}}}$ and $R_{2b^{t\bar{t}} \rightarrow 4b^{t\bar{t}}}$ reweightings did not give good closure. The hyperparameters were re-optimized to give a model with 3 hidden layers of 50, 25, and 10 neurons with batch normalization [124] between the layers. The closures on the reweighting variables (and m_{HH}) are shown in Figure F.15 and Figure F.16 for these two reweightings – and demonstrates that the individual reweightings going into Eq. F.5 were optimized satisfactorily.

However, when we applied the weights from Eq. F.5, the resulting X_{W_t} distribution had an (unphysical) discontinuous QCD template in the $X_{W_t} < 1.5$ region, with the following issues:

- When $R_{2b^{t\bar{t}} \rightarrow 4b^{t\bar{t}}}$ was slightly larger than 1, there was large weight amplification as the denominator went to zero.
- We also had some unphysically large weights from the $R_{2b^{t\bar{t}} \rightarrow 4b^{t\bar{t}}}$ reweighting, suggesting that a region of low support for evaluating on the 2b data.
- Sometimes the subtraction in the numerator caused a negative weight, which could become amplified by a small denominator.

Figure F.15: $R_{2b^{t\bar{t}} \rightarrow 4b^{t\bar{t}}}$: The MC based $2b \rightarrow 4b$ $t\bar{t}$ reweighting.Figure F.16: $R_{2b^{t\bar{t}} \rightarrow 4b^{t\bar{t}}}$: Reweighting $2b$ data $\rightarrow 2b$ $t\bar{t}$.

To fix the unphysical QCD template issue, we removed the unphysical weights by cutting out the events with a weight value more than 10σ away from the mean. The resulting pre-fit templates for all of the years are shown in the left-hand column of Figure F.17, a reasonable template for the pure QCD distribution.

We then fit the 2 component normalizations to the X_{Wt} shape inside in CR1 (results in Table F.2), and checked the post-fit agreement in the left-hand column of Figure F.17. The modeling is always better in the $X_{Wt} < 1.5$ region, although the modeling also appears improved for the $X_{Wt} > 1.5$ region. Testing the setup with all the 100 replicas and including these error bars would be needed to further validate the performance.

	2016	2017	2018
$\alpha_{t\bar{t}}$	1.27 ± 0.11	1.84 ± 0.09	0.66 ± 0.08
α_{QCD}	0.950 ± 0.018	0.913 ± 0.014	1.012 ± 0.013

Table F.2: Fitted normalizations for the $t\bar{t}$ and QCD templates with the pure QCD reweighting. The quoted uncertainties are from the 4b X_{Wt} shape fit.

Although preliminary, these results are already interesting as a first proof-of-principle for how to account for separate background components, useful both for treating either $t\bar{t}$ or single Higgs backgrounds independently. But although the pure QCD reweighting was mathematically elegant, implementing it in practice wasn't trivial as we had to apply these ad-hoc cuts to the final weights to get a meaningful template. In Chapter sec:pure-qcd-flow, we show how $t\bar{t}$ can also be modeled separately with direct density estimation methods – an avenue that may be more promising for pursuing in the future.

After Run 3, our dataset size will *double*, making the intricacies of how we treat the separate components of our background estimate even more relevant. As we transition away from setting limits and start trying to extract the HH signal, it will become even more important to separately constraining each component of the background. This chapter focused on $t\bar{t}$ background, and even though it is only 10% of our background, it is 100x as large as our HH signal. The single Higgs background will also become more important as our single Higgs yield in the 4b SR is 3x the SM HH signal. The $bb\gamma\gamma$ analysis already includes the single Higgs background in their background modeling for the limit extraction, but the same methods from Chapter F.3.1 can be used to model the single Higgs background for 4b, and will be helpful for HHarmonization with the other channels in the quest to extract the HH signal.

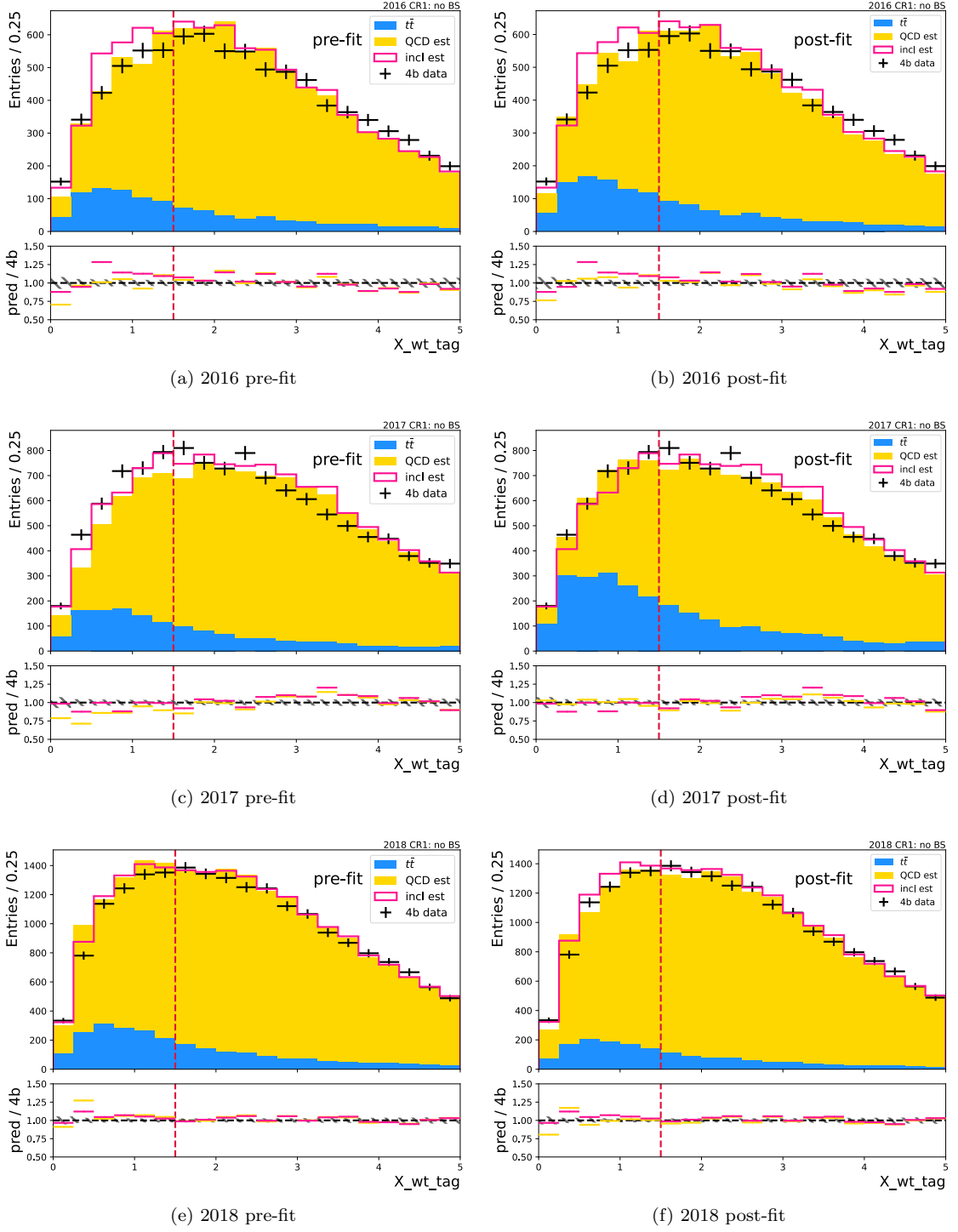


Figure F.17: X_{Wt} distributions in the 4b CR1 with the pre-fit (left) and post-fit (right) plots, compared to the 4b CR1 data. The uncertainties in the subpanels of the right column are from the 4b X_{Wt} shape fit.

Appendix G

Flows optimization studies

G.1 Real NVP

One of the simplest types of flow architectures is the Real-NVP [187]. To keep the Jacobian inverse computation tractable, we take turns transforming the variables, so the Jacobian becomes block diagonal. This reduces the inverse computation from $\mathcal{O}(n^3)$ to $\mathcal{O}(n^2)$ (where n is the dimension of the modeling variables). To ensure that the matrix stays block diagonal, each f_i step of the flow transforms half of the variable dimensions at a time.

Figure G.1 shows a schematic for a two variable flow modeling m_{hh} and $|\cos \Theta^*|$ (where Θ^* is the polar angle of one of the HCs in the HH rest frame). Starting from a 2d Gaussian density, each f_i step of the flow successively modifies the vector z to predict the HH kinematics. For example, in the first step of the flow (f_0) the first dimension of the flow z_0 is transformed while z_1 stays fixed, while in the next step of the flow (f_2), the z_1 dimension transforms while the z_0 is fixed.

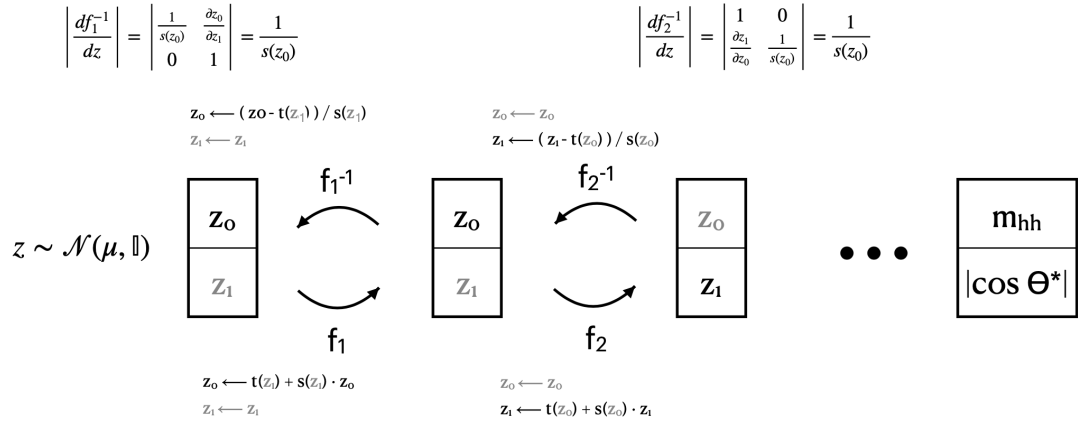


Figure G.1: Graphic for explaining how a Real-NVP works for a 2 variable flow.

To preserve the bijection, affine transformations are used. The scale (s) and shift (t) of the affine transformation depend on the other (non-transforming) variables and are predicted by small NNs. The NN parameters are optimized by maximizing the likelihood of the data, $p_y(y)$. To ensure the positivity of s , the NN instead predicts $\log s$ and then exponentiates the output.

Since the variables go one-at-a-time, the inverse computation remains tractable, i.e, when z_1 is known, it's trivial to solve for f_0^{-1} as $z_0 \rightarrow (z_1 - t(z_1)) / s(z_1)$.

The Real-NVP is equally fast for density estimation (i.e, model training) and sampling (or predicting the kinematics distributions). Our preliminary 4b background optimizations started from this architecture. To solve the interpolation problem, we predict the conditional probability distribution (also shown in [187]) by passing $(m_{h1}), m_{h2})$ as additional variables to the scale and shift networks. This still preserves the nice properties of the bijection for each of the flow steps while the Jacobians remain tractable with the modification to the math shown in Figure G.2.

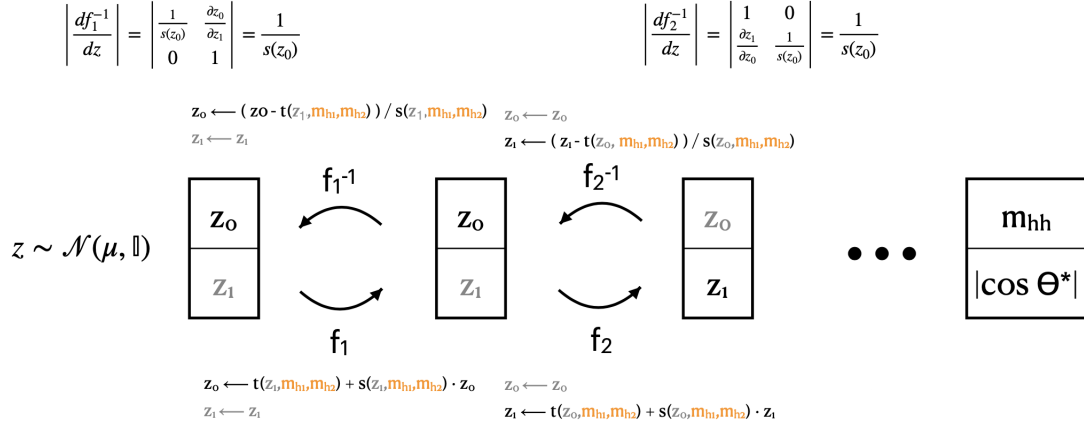


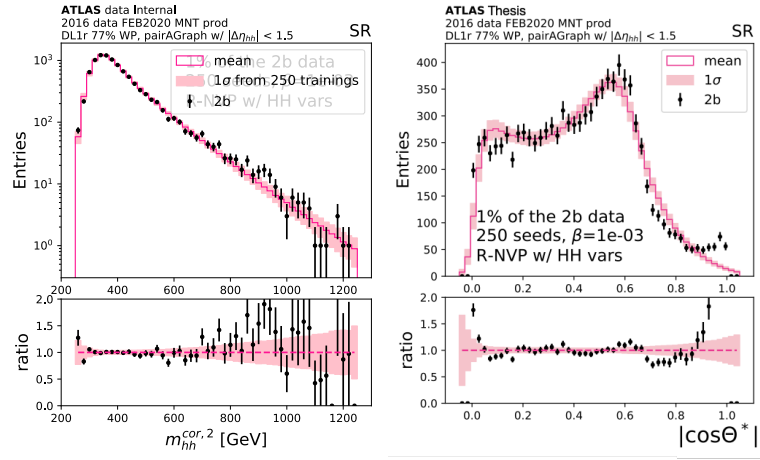
Figure G.2: The extension of the Real-NVP for a conditional 2 variable flow.

G.2 flow optimizations

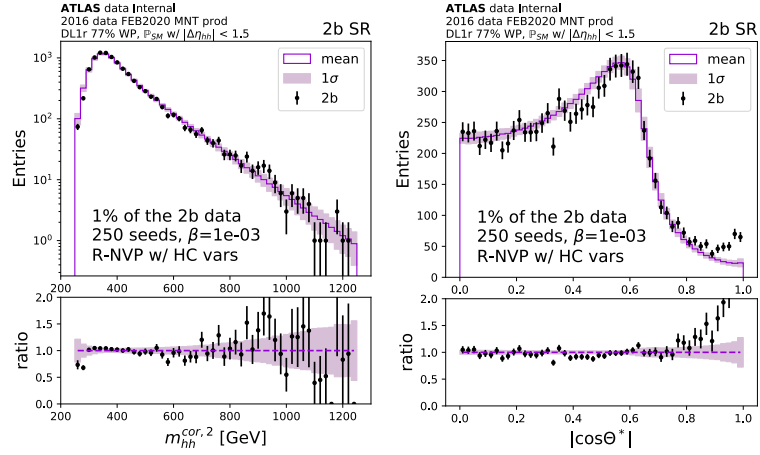
We carried out most of the optimization studies with the analysis configuration with pairAGraph for the HC pairing and massplane definition Appendix ???. Also, we did most of the optimization modeling a (downsampled) portion of the 2b massplane to avoid biasing ourselves in looking at the 4b SR before the non-resonant analysis had unblinded.

G.2.1 Modelling variables

Figure G.3 shows a comparison of training on the m_{HH} and $|\cos \Theta^*|$ directly (top row: prediction in pink) versus training on the HC 3-vectors (bottom row: prediction in purple). We trained to model



(a) Real NVP directly modeling HH variables



(b) Real NVP modelling the Higgs 3-vectors to predict the HH variables

Figure G.3: Comparison of modeling HH kinematics versus the HC 3-vectors.

a corrected version of m_{HH} :

$$m_{HH}^{cor,2} = m_{HH} - m_{H1} - m_{H2} + 250 \text{ GeV}$$

. This variable is always larger than 250 GeV, so we trained the flow to model $\log(m_{HH}^{cor,2} - 250)$. Although the flow had comparable results for the $m_{HH}^{cor,2}$ whether it was modeled directly or trained with the HCs, $|\cos\Theta^*|$ was easier to model when we reconstructed the HCs since this allowed us to learn a distribution that was always bounded between $[0,1]$.

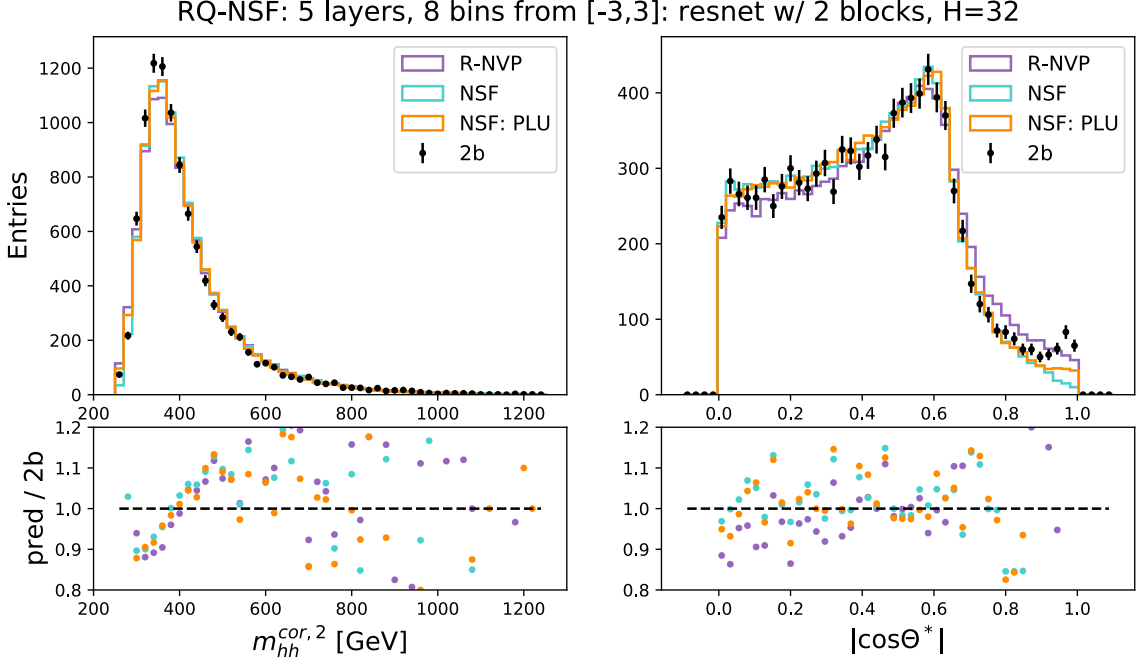


Figure G.4

G.2.2 Choice of flow model

Figure G.4 shows our comparison (with just a single training) for the m_{HH} and $|\cos \Theta^*|$ modeling for the Real-NVP architecture (purple), the NSF without mixing-up the variables between the coupling layers (turquoise), and the NSF with the $W = PLU$ generalized permutations¹ (orange). Since the NSF with the $W = PLU$ layers for variable mixing seemed to be the best model we went ahead to do the hyperparameter optimization with this setup

G.2.3 Hyperparameter optimization

The rational quadratic neural spline flow paper considered modeling a range of applications from the UCI dataset, so we used a range of hyperparameters that seemed to give good performance on the datasets with a similar number of training examples (see Table 4 of [173]).

We considered $B=3$, i.e, the rational quadratic spline transform is defined for the variable from $(-3,3)$, and identity outside of this range. To keep the inputs in the range where the NSF will have a meaningful transformation, a batch normalization layer was included after each of the flow steps. We didn't consider annealing the learning rate or gradient clipping even though this was used for the experiments in [173] since we had good enough performance for our application before trying to

¹Described in Eq. 13.11

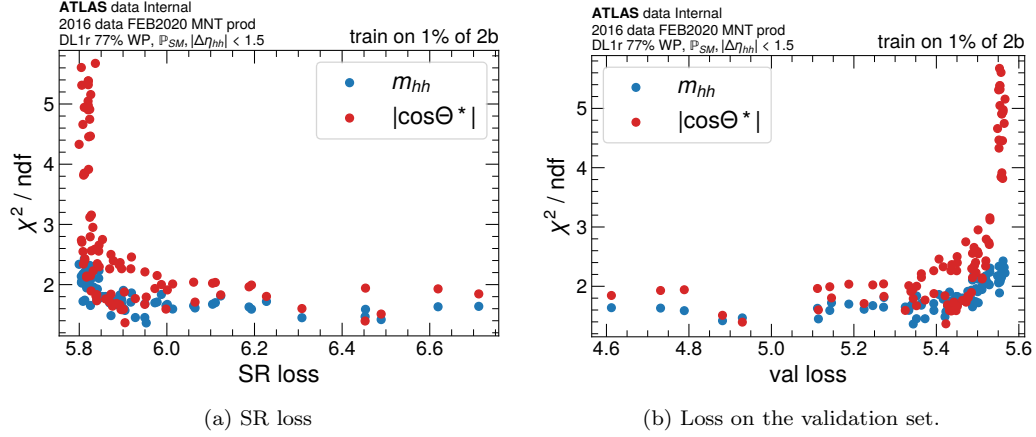


Figure G.5: To choose the optimal model we found that minimizing the validation loss was a better metric than minimizing the SR loss.

include this. The choices for hyperparameters that I was considering were:

Parameter range	parameter description
$L = [5, 10]$	number of layers (or flow steps)
$K = [4, 8]$	number of knots in each spline transformation
$\alpha = 3\text{e-}4, 5\text{e-}4, 1\text{e-}3$	learning rate (for adam optimizer)
$p = 0, 0.1, 0.2$	dropout fraction
$\beta = 0, 1\text{e-}6, 1\text{e-}5, 1\text{e-}4, 1\text{e-}3$	L2 regularization weight.
num_blocks = 1, 2	The number of blocks for the res net predicting the spline parameters
$H = 16, 32, 64$	hidden dimensions of the NNs.

Table G.1: Hyperparameters ranges considered for the random search.

We did a random search over the hyperparameters to train 90 different configurations, where each configuration choice was trained 25 times. We looked at the χ^2 of the m_{HH} and $|\cos\Theta^*|$ 2b SR histograms to determine which were the best-performing models. Figure G.5 shows the correlation between these χ^2 s and the validation loss (left) and the SR loss (right). The loss on the validation set was a better indicator of the interpolation closure than the negative log likelihood in the SR. We found that the more complex models had a higher loss (because they were predicting that these SR events were anomalous), but actually produced better SR samples, consistent with the findings from [188] Figure G.6 shows a subset of the models considered in the hyperparameter scan, and indicates that models with lower training and validation loss tended to have a higher SR loss. We used one of these models for the suite of background validation experiments and to look at the 4b SR model.

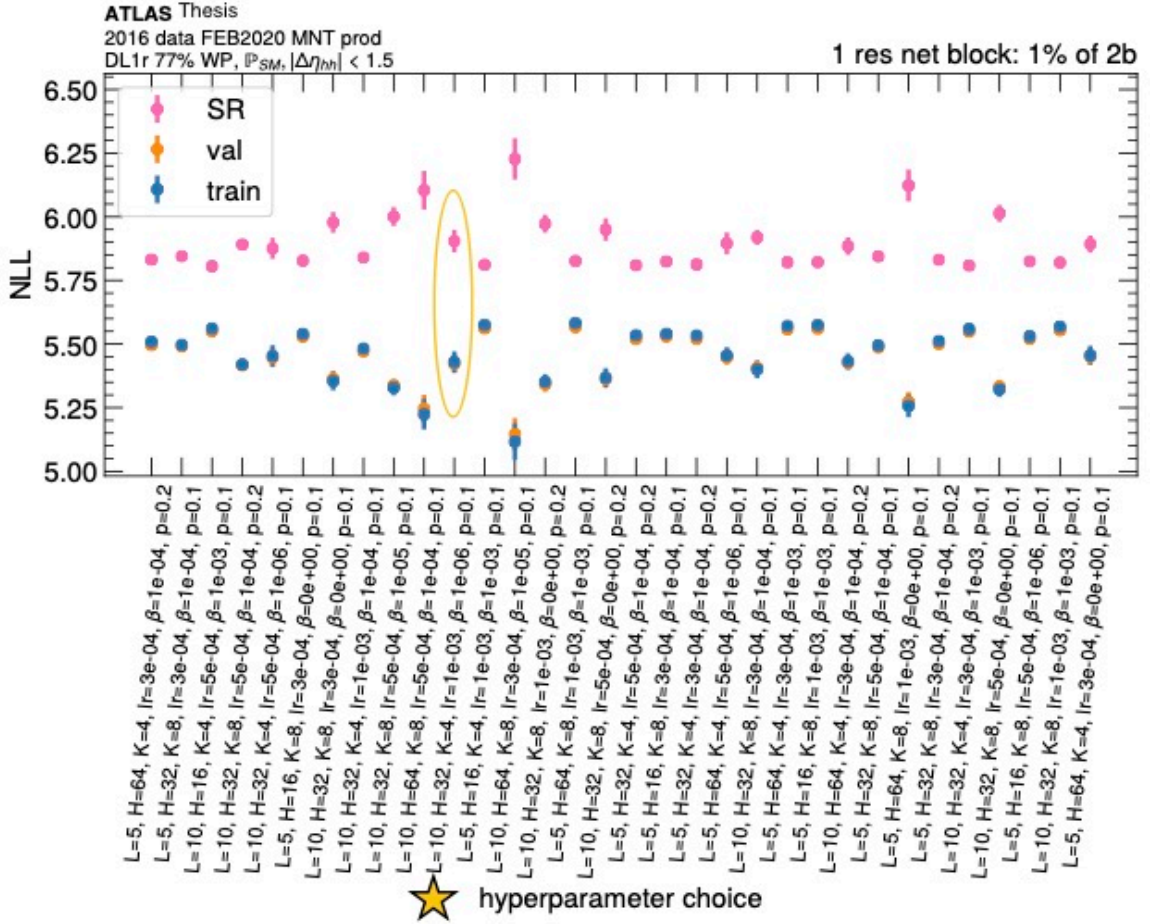


Figure G.6: Illustration of a subset of the models considered in the hyperparameter scan. The orange and blue show the training and validation loss in the interpolation bounding box (with the SR), and the pink shows the negative log likelihood of the SR data. Each model configuration has been trained 25 times, and the mean and standard deviation are plotted. The star (and yellow circle) shows the model configuration that we used for the results in Chapter A.1.

Bibliography

- [1] *Performance of 2019 recommendations of ATLAS Flavor Tagging algorithms with Variable Radius track jets*, <http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2019-006/>, 2019 (**backrefpages** v, 90, 91).
- [2] *Flavour-tagging efficiency corrections for the 2019 ATLAS PFlow jet b-taggers with the full LHC Run II dataset*, <http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2020-001/>, 2020 (**backrefpages** v, 93, 96).
- [3] A. Collaboration, *ATLAS flavour-tagging algorithms for the LHC Run 2 pp collision dataset*, The European Physical Journal C **83** (2023), URL: <https://arxiv.org/abs/2211.16345> (**backrefpages** v, 82, 86, 88).
- [4] *Light flavour mistag calibration of the 2019 ATLAS particle-flow jets b-taggers using the Z+jets Negative Tag method*, <http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2021-002/>, 2021 (**backrefpages** v, 94–96).
- [5] ATLAS Collaboration, *Deep Sets based Neural Networks for Impact Parameter Flavour Tagging in ATLAS*, ATL-PHYS-PUB-2020-014, 2020, URL: <https://cds.cern.ch/record/2718948> (**backrefpages** v, 70, 72, 95, 96, 100–103, 105–108).
- [6] N. Hartman, R. Teixeira de Lima, M. Kagan, on behalf of the ATLAS Collaboration, *Deep Sets for Flavor Tagging at the ATLAS Experiment*, [Proceedings of the 2020 Connecting The Dots Workshop \(\)](#), URL: <https://zenodo.org/record/4088760#.YUB1op4zb0p> (**backrefpage** v).
- [7] ATLAS Collaboration, *Search for non-resonant pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using 126 fb^{-1} of pp collision data at $\sqrt{s} = 13\text{ TeV}$ with the ATLAS detector.*, ATLAS-CONF-2022-035, 2022, URL: <https://cds.cern.ch/record/2811390> (**backrefpages** vi, 181, 234).
- [8] ATLAS Collaboration, *Search for nonresonant pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state in pp collisions at $\sqrt{s} = 13\text{ TeV}$ with the ATLAS detector*, *Phys. Rev. D* **108** (2023) [page 052003](#), arXiv: [2301.03212](https://arxiv.org/abs/2301.03212) [[hep-ex](#)] (**backrefpage** vi).
- [9] M. Peskin, *Concepts of Elementary Particle Physics*, Oxford University Press, 2019 (**backrefpages** viii, 7).

- [10] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012) [page 1](#), arXiv: [1207.7214 \[hep-ex\]](#) (**backrefpages 1, 21, 27, 117**).
- [11] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Phys. Lett. B* **716** (2012) [page 30](#), arXiv: [1207.7235 \[hep-ex\]](#) (**backrefpages 1, 21, 27, 117**).
- [12] *Standard Model*, wikipedia, URL: https://en.wikipedia.org/wiki/Standard_Model (**backrefpage 2**).
- [13] Y. Fukuda **and others**, *Evidence for Oscillation of Atmospheric Neutrinos*, *Phys. Rev. Lett.* **81** (8 1998) [page 1562](#), URL: <https://link.aps.org/doi/10.1103/PhysRevLett.81.1562> (**backrefpage 2**).
- [14] The KATRIN Collaboration, *Direct neutrino-mass measurement with sub-electronvolt sensitivity*, *Nature Physics* **18** (2022) [page 160](#) (**backrefpage 2**).
- [15] Muon g_2 Collaboration, *Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm*, *Physical Review Letters* **126** (2021), URL: <https://doi.org/10.1103/2Fphysrevlett.126.141801> (**backrefpage 2**).
- [16] CDF Collaboration, *High-precision measurement of the W boson mass with the CDF II detector*, *Science* **376** (2022), URL: <https://doi.org/10.1126/science.abk1781> (**backrefpage 2**).
- [17] Melissa Quinnan, *Standard model four-top quark production in the all-hadronic final state in proton-proton collisions at 13 TeV with the CMS experiment*, 2022 (**backrefpage 6**).
- [18] A. Larkoski, *Elementary Particle Physics: An Intuitive Introduction*, Cambridge University Press, 2019 (**backrefpages 7, 20**).
- [19] C. S. Wu, E. Amber, R. W. Hayward, D. D. Hoppes **and** R. P. Hudson, *Experimental Test of parity conservation in beta decay*, *Physical Review* **105** (1957) (**backrefpage 7**).
- [20] wikipedia, URL: https://commons.wikimedia.org/wiki/File:Mexican_hat_potential_polar_with_details.svg (**backrefpage 8**).
- [21] CMS Collaboration, *Measurement and QCD analysis of double-differential inclusive jet cross sections in pp collisions at $\sqrt{s} = 8$ TeV and cross section ratios to 2.76 and 7 TeV*, *JHEP* **03** (2017) [page 156](#), arXiv: [1609.05331 \[hep-ex\]](#) (**backrefpage 16**).
- [22] A. J. Larkoski, *An Unorthodox Introduction to QCD*, 2017, URL: <https://arxiv.org/abs/1709.06195> (**backrefpage 17**).
- [23] A. J. Larkoski, I. Mout **and** B. Nachman, *Jet substructure at the Large Hadron Collider: A review of recent advances in theory and machine learning*, *Physics Reports* **841** (2020) [page 1](#), URL: <https://doi.org/10.1016/j.physrep.2019.11.001> (**backrefpages 17, 59**).

- [24] Tara Sears, *The Particle World: an introduction to particle physics*, CERN Summer Student Lectures, 2014 ([backrefpage 17](#)).
- [25] A. Cho, *Mass of the Common Quark Finally Nailed Down*, *Science* (2010), URL: <https://www.science.org/content/article/mass-common-quark-finally-nailed-down> ([backrefpage 18](#)).
- [26] *Martin-Stirling-Thorne-Watt Parton Distribution Functions*, URL: <https://mstwpdf.hepforge.org> ([backrefpage 18](#)).
- [27] S. Höche, *Introduction to parton-shower event generators*, 2014, URL: <https://arxiv.org/abs/1411.4085> ([backrefpage 19](#)).
- [28] J. Ellis, M. K. Gaillard and D. V. Nanopoulos, *A Historical Profile of the Higgs Boson*, (2012), URL: <https://arxiv.org/abs/1201.6045> ([backrefpage 20](#)).
- [29] C. T. Potter and others, *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group*, en, 2013, URL: <http://cds.cern.ch/record/1559921> ([backrefpage 20](#)).
- [30] ATLAS Collaboration, (2020), URL: <https://arxiv.org/abs/2004.003447> ([backrefpage 21](#)).
- [31] ATLAS Collaboration, *Measurement of the properties of Higgs boson production at $\sqrt{s} = 13$ TeV in the $H \rightarrow$ channel using 139 fb^{-1} of pp collision data with the ATLAS experiment*, (2022), URL: <https://arxiv.org/abs/2207.00348> ([backrefpage 21](#)).
- [32] *A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery*, *Nature* **607** (2022) [page 52](#), URL: <https://doi.org/10.1038%2Fs41586-022-04893-w> ([backrefpages 21–23](#)).
- [33] ATLAS Collaboration, *Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector*, *Eur. Phys. J. C* **75** (2015) [page 476](#), arXiv: [1506.05669 \[hep-ex\]](#) ([backrefpage 23](#)).
- [34] F. Caola and K. Melnikov, *Constraining the Higgs boson width with $Z Z$ production at the LHC*, *Physical Review D* **88** (2013), URL: <https://arxiv.org/abs/1307.4935> ([backrefpage 23](#)).
- [35] J. Campbell, R. Ellis and C. Williams, *Bounding the Higgs width at the LHC using full analytic results for $gg \rightarrow e^- e^+ \mu^- \mu^+$* , *Journal of High Energy Physics* **60** (2014) ([backrefpage 23](#)).
- [36] R. C. L. D. Sa, *Higgs total width at the ATLAS experiment*, LHC Seminar (2022), URL: <https://indico.cern.ch/event/1187940/> ([backrefpage 23](#)).
- [37] L. Evans and P. Bryant, *LHC Machine*, *JINST* **3** (2008) [S08001](#) ([backrefpage 24](#)).
- [38] CERN, *The Large Hadron Collider*, URL: <https://home.cern/science/accelerators/large-hadron-collider> ([backrefpage 24](#)).

- [39] CERN, *LHC The guide FAQ*, (), URL: <https://home.cern/resources/brochure/knowledge-sharing/lhc-facts-and-figures> (backrefpage 25).
- [40] F. Sauerburger, *LHC cross section plot*, (), URL: <https://lhc-xsecs.org/> (backrefpage 26).
- [41] *The HL LHC project*, (), URL: <https://hilumilhc.web.cern.ch/content/hl-lhc-project> (backrefpage 27).
- [42] ATLAS Collaboration, *Public ATLAS Luminosity Results for Run-2 of the LHC*, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2\#Pileup_Interactions_and_Data_Tak (backrefpages 27, 28).
- [43] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08003 (backrefpages 29, 32, 35, 37, 41–44).
- [44] J. Pequeno and P. Schaffner, *How ATLAS detects particles: diagram of particle paths in the detector*, <https://cds.cern.ch/record/1505342>, 2013 (backrefpage 30).
- [45] wikipedia, URL: <https://en.wikipedia.org/wiki/Pseudorapidity> (backrefpage 31).
- [46] ATLAS Collaboration, *Software Performance of the ATLAS Track Reconstruction for LHC Run 3*, ATL-PHYS-PUB-2021-012, 2021, URL: <https://cds.cern.ch/record/2766886/> (backrefpage 32).
- [47] G. L. Aad, M. Ackers, F. A. Aleppo and et al., *ATLAS Pixel Detector Electronics and Sensors*, *Jinst* **3** (2008) P07007 (backrefpages 32–34).
- [48] Jennet Dickenson, *ATLAS Measurements of the Higgs Boson Coupling to the Top Quark in the Higgs to Diphoton Decay Channel*, CERN-THESIS-2020-043, 2020 (backrefpage 32).
- [49] S. Zenz, *How a Pixel Detector Works.*, Imperial College London Blog, URL: <http://www.quantumdiaries.org/2008/07/25/how-a-pixel-detector-works> (backrefpage 33).
- [50] S. Stapnes, *Detector Challenges at the LHC*, *Nature* **448** (2007) (backrefpage 34).
- [51] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, ATLAS-TDR-19, 2010, URL: <https://cds.cern.ch/record/1291633> (backrefpage 34).
- [52] B. Abbott and others, *Production and integration of the ATLAS Insertable B-Layer*, *JINST* **13** (2018) T05008, arXiv: 1803.00844 [physics.ins-det] (backrefpage 34).
- [53] A. Collaboration, *Impact Parameter Resolution*, <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/IDTR-2015-007/>, 2015 (backrefpage 35).
- [54] A. Romaniouk, *ATLAS TRT (Transition Radiation Tracker)*, 2019, URL: <http://particle.mephi.ru/en/research/atlas/trt/> (backrefpage 36).
- [55] R. Fernow, *Introduction to experimental particle physics*, 1990 (backrefpage 38).
- [56] J. E. Brau, J. A. Jaros and H. Ma, *Advances in Calorimetry*, PH-EP-Internal-Technical-Note-2010-013 (2010) (backrefpage 38).

- [57] M. Gupta, *Calculation of Radiation Length in materials.*, Annual Reviews **41** (1991) (**backrefpage 38**).
- [58] E. Halkiadakis, *Proceedings for TASI 2009 Summer School on "Physics of the Large and the Small": Introduction to the LHC experiments*, 2010, URL: <https://arxiv.org/abs/1004.5564> (**backrefpage 38**).
- [59] Marco Valente, *Supersymmetric Beasts and Where to Find Them: From Novel Hadronic Reconstruction Methods to Search Results in Large Jet Multiplicity Final States at the ATLAS Experiment*, CERN-THESIS-2020-047, 2020 (**backrefpage 39**).
- [60] A. Collaboration, *Technical Design Report: Liquid Argon, Ch 1 and Ch 6*, (1996) (**backrefpage 40**).
- [61] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, *Eur. Phys. J. C* **77** (2017) **page 490**, arXiv: 1603.02934 [[hep-ex](#)] (**backrefpages 40, 55**).
- [62] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, *Eur. Phys. J. C* **77** (2017) **page 466**, arXiv: 1703.10485 [[hep-ex](#)] (**backrefpages 41, 55, 56**).
- [63] E. Govorkova **and others**, *LHC physics dataset for unsupervised New Physics detection at 40 MHz*, (2021), URL: <https://arxiv.org/abs/2107.02157> (**backrefpage 44**).
- [64] ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton–proton collisions at $\sqrt{s} = 7$ TeV*, *Eur. Phys. J. C* **73** (2013) **page 2304**, arXiv: 1112.6426 [[hep-ex](#)] (**backrefpage 45**).
- [65] R. E. Kalman, *A new approach to linear filtering and prediction problems*, Transactions of the ASME Journal of Basic Engineering **82** (1960) **page 35** (**backrefpage 46**).
- [66] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*, Nuclear Instrumentation Methods **A262** (1987) **page 444** (**backrefpages 46, 51**).
- [67] Andy Salzburger, *Track Simulation and Reconstruction in the ATLAS Experiment*, 2008, URL: http://physik.uibk.ac.at/hephy/theses/diss_as.pdf (**backrefpage 46**).
- [68] Nicola Giacinto Piacquadio, *Identification of b-jets and investigation of the discovery potential of a Higgs boson in the $WH \rightarrow \ell \nu b \bar{b}$ channel with the ATLAS experiment*, CERN-THESIS-2010-027, 2010 (**backrefpages 47, 50, 51, 72, 79, 83**).
- [69] H. Gray, *Track reconstruction in the ATLAS experiment*, Experimental Particle and Astroparticle Seminar Zurich, 2016, URL: <https://indico.cern.ch/event/504284/> (**backrefpage 48**).
- [70] T. Cornelissen and M. Elsing and I. Gavrilenko and S. Fleischmann and W. Liebig and E. Moyse and A. Salzburger, *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)*, ATL-SOFT-PUB-2007-007, ATL-COM-SOFT-2007-002, 2007, URL: <https://cds.cern.ch/record/1020106> (**backrefpage 48**).

- [71] ATLAS Collaboration, *A neural network clustering algorithm for the ATLAS silicon pixel detector*, *JINST* **9** (2014) P09009, arXiv: 1406.7690 [hep-ex] (backrefpages 48, 49).
- [72] P. Billoir and S. Qian, *Fast vertex fitting with a local parametrization of tracks*, *Nuclear Instrumentation Methods* **A311** (1992) page 139 (backrefpage 51).
- [73] ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC*, *Eur. Phys. J. C* **77** (2017) page 332, arXiv: 1611.10235 [hep-ex] (backrefpages 52, 53).
- [74] T. Robertson and J. D. Cryer, *An Iterative Procedure for Estimating the Mode*, *Journal of the American Statistical Association* **69** (1974) page 1012, eprint: <https://www.tandfonline.com/doi/pdf/10.1080/01621459.1974.10480246>, URL: <https://www.tandfonline.com/doi/abs/10.1080/01621459.1974.10480246> (backrefpage 53).
- [75] W. Waltenberger, R. Frühwirth and P. Vanlaer, *Adaptive vertex fitting*, *J. Phys. G* **34** (2007) N343 (backrefpage 53).
- [76] ATLAS Collaboration, *ATLAS event at 13 TeV – First stable beam, 3 June 2015 – run: 266904, evt: 25884805*, <https://cds.cern.ch/record/2022202>, 2015 (backrefpage 53).
- [77] M. Cacciari, G. P. Salam and G. Soyez, *The anti- k_t jet clustering algorithm*, *JHEP* **04** (2008) page 063, arXiv: 0802.1189 [hep-ph] (backrefpage 54).
- [78] ATLAS Collaboration, *Jet energy scale and resolution measured in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Eur. Phys. J. C* **81** (2021) page 689, arXiv: arXiv:2007.02645 [hep-ex] (backrefpages 56, 57).
- [79] K. Pascal, *JES / JER for JDM*, 2018, URL: <https://indico.cern.ch/event/777980/contributions/3236356/subcontributions/271401/attachments/1780062/2896765/JESJER.pdf> (backrefpage 57).
- [80] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, ATLAS-CONF-2014-018, 2014, URL: <https://cds.cern.ch/record/1700870> (backrefpage 58).
- [81] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, *Eur. Phys. J. C* **76** (2016) page 581, arXiv: 1510.03823 [hep-ex] (backrefpage 58).
- [82] D. Krohn, J. Thaler and L.-T. Wang, *Jet trimming*, *Journal of High Energy Physics* **2010** (2010), URL: <https://doi.org/10.1007/jhep02%282010%29084> (backrefpage 58).
- [83] ATLAS Collaboration, *Variable Radius, Exclusive- k_T , and Center-of-Mass Subjet Reconstruction for Higgs($\rightarrow b\bar{b}$) Tagging in ATLAS*, ATL-PHYS-PUB-2017-010, 2017, URL: <https://cds.cern.ch/record/2268678> (backrefpages 58–60).

- [84] ATLAS Collaboration, *Boosted Higgs ($\rightarrow b\bar{b}$) Boson Identification with the ATLAS Detector at $\sqrt{s} = 13$ TeV*, ATLAS-CONF-2016-039, 2016, URL: <https://cds.cern.ch/record/2206038> (backrefpages 59, 68).
- [85] D. Krohn, J. Thaler and L.-T. Wang, *Jets with variable R* , *Journal of High Energy Physics* **2009** (2009) page 059, URL: <https://doi.org/10.1088/1126-6708/2009/06/059> (backrefpage 59).
- [86] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015-2017 LHC proton-proton collision data*, (2019), arXiv: 1908.00005 [hep-ex] (backrefpage 61).
- [87] A. Collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **81** (2021) page 578, URL: <https://doi.org/10.1140/epjc/s10052-021-09233-2> (backrefpage 61).
- [88] I. Neutelings, *Hadronically decayed tau jet*, https://tikz.net/jet_tau/, 2021 (backrefpage 62).
- [89] ATLAS Collaboration, *Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment*, ATL-PHYS-PUB-2019-033, 2019, URL: <https://cds.cern.ch/record/2688062> (backrefpage 62).
- [90] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **78** (2018) page 903, arXiv: 1802.08168 [hep-ex] (backrefpage 63).
- [91] ATLAS Collaboration, *Configuration and performance of the ATLAS b -jet triggers in Run 2*, *Eur. Phys. J. C* **81** (2021) page 1087, arXiv: 2106.03584 [hep-ex] (backrefpages 65, 132, 133).
- [92] ATLAS Collaboration, *Performance of the ATLAS detector using first collision data*, *JHEP* **09** (2010) page 056, arXiv: 1005.5254 [hep-ex] (backrefpage 66).
- [93] ATLAS Collaboration, *Identification of Jets Containing b -Hadrons with Recurrent Neural Networks at the ATLAS Experiment*, ATL-PHYS-PUB-2017-003, 2017, URL: <https://cds.cern.ch/record/2255226> (backrefpages 67, 74–76, 78, 95).
- [94] S. Frixione, P. Nason and G. Ridolfi, *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, *JHEP* **09** (2007) page 126, arXiv: 0707.3088 [hep-ph] (backrefpage 68).
- [95] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, *JHEP* **11** (2004) page 040, arXiv: hep-ph/0409146 (backrefpage 68).
- [96] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, *JHEP* **11** (2007) page 070, arXiv: 0709.2092 [hep-ph] (backrefpage 68).

- [97] S. Alioli, P. Nason, C. Oleari **and** E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010) [page 043](#), arXiv: [1002.2581 \[hep-ph\]](#) ([backrefpage 68](#)).
- [98] T. Sjöstrand **and others**, *An introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) [page 159](#), arXiv: [1410.3012 \[hep-ph\]](#) ([backrefpage 68](#)).
- [99] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014, URL: <https://cds.cern.ch/record/1966419> ([backrefpage 68](#)).
- [100] D. J. Lange, *The EvtGen particle decay simulation package*, *Nucl. Instrum. Meth. A* **462** (2001) [page 152](#) ([backrefpage 68](#)).
- [101] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, *Eur. Phys. J. C* **70** (2010) [page 823](#), arXiv: [1005.4568 \[physics.ins-det\]](#) ([backrefpage 68](#)).
- [102] S. Agostinelli **and others**, GEANT4 – a simulation toolkit, *Nucl. Instrum. Meth. A* **506** (2003) [page 250](#) ([backrefpage 68](#)).
- [103] A. Chisholm, *Limiting factors in ATLAS flavour tagging calibration measurements*, https://indico.cern.ch/event/895505/contributions/3779824/attachments/2023559/3384288/ASC_FTAGWS20_CalibLimits.pdf, 2020 ([backrefpage 71](#)).
- [104] S. Ermon, *Probabilistic Graphical Models, Lecture 5*, Winter 2019 ([backrefpage 73](#)).
- [105] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV*, (2019), arXiv: [1907.05120 \[hep-ex\]](#) ([backrefpages 73, 93](#)).
- [106] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*, ATL-PHYS-PUB-2017-013, 2017, URL: <https://cds.cern.ch/record/2273281> ([backrefpages 74, 84, 136](#)).
- [107] S. Hochreiter **and** J. Schmidhuber, *Long Short-Term Memory*, *Neural Computation* **9** **8** (1997) ([backrefpage 75](#)).
- [108] F.-F. Li, J. Johnson **and** S. Yeung, *Lecture 10: Recurrent Neural Networks*, Spring 2018 ([backrefpage 75](#)).
- [109] S. Hochreiter **and** J. Schmidhuber, *Long short-term memory*, *Neural computation* **9** (1997) [page 1735](#) ([backrefpage 75](#)).
- [110] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever **and** R. Salakhutdinov, *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*, 2014 ([backrefpages 76, 198](#)).
- [111] T. Huffman, C. Jackson **and** J. Tseng, *Tagging b-quarks at extreme energies without tracks*, *Journal of Physics G: Nuclear and Particle Physics* **43** (2016) [page 085001](#), URL: <https://doi.org/10.1088/0954-3899/43/8/085001> ([backrefpage 77](#)).

- [112] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, 2014, arXiv: [1412.6980 \[cs.LG\]](#) (backrefpages 78, 87, 99, 155).
- [113] ATLAS Collaboration, *Secondary vertex finding for jet flavour identification with the ATLAS detector*, ATL-PHYS-PUB-2017-011, 2017, URL: <https://cds.cern.ch/record/2270366> (backrefpage 79).
- [114] ATLAS Collaboration, *Topological b-hadron decay reconstruction and identification of b-jets with the JetFitter package in the ATLAS experiment at the LHC*, ATL-PHYS-PUB-2018-025, 2018, URL: <https://cds.cern.ch/record/2645405> (backrefpage 83).
- [115] A. Collaboration, *Expected performance of the 2019 ATLAS b-taggers*, <http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2019-005/>, 2019 (backrefpages 88, 89, 135).
- [116] ATLAS Collaboration, *Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *JHEP* **01** (2019) page 030, arXiv: [1804.06174 \[hep-ex\]](#) (backrefpages 92, 150, 154, 155, 165, 188, 246).
- [117] ATLAS Collaboration, *Search for resonant pair production of Higgs bosons pairs in the $b\bar{b}b\bar{b}$ final state using pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Rev. D* **105** (2022) page 092002, arXiv: [2202.07288 \[hep-ex\]](#) (backrefpages 92, 150, 165, 191).
- [118] ATLAS Collaboration, *Simulation-based extrapolation of b-tagging calibrations towards high transverse momenta*, ATL-PHYS-PUB-2021-003, 2021, URL: <https://cds.cern.ch/record/2753444> (backrefpage 94).
- [119] A. Collaboration, *Measurement of the c-jet mistagging efficiency in $t\bar{t}$ events using pp collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS detector*, *The European Physical Journal C* **82** (2022) page 95, arXiv: [2109.10627 \[hep-ex\]](#), URL: <https://doi.org/10.1140/epjc/s10052-021-09843-w> (backrefpage 94).
- [120] ATLAS Collaboration, *Calibration of light-flavour b-jet mistagging rates using ATLAS proton–proton collision data at $\sqrt{s} = 13$ TeV*, ATL-CONF-2018-006, 2018, URL: <https://cds.cern.ch/record/2314418> (backrefpage 94).
- [121] M. Zaheer and others, *Deep Sets*, CoRR (2017), URL: <http://arxiv.org/abs/1703.06114> (backrefpage 97).
- [122] P. T. Komiske, E. M. Metodiev and J. Thaler, *Energy flow networks: Deep Sets for particle jets*, *JHEP* **01** (2019) page 121, URL: [http://dx.doi.org/10.1007/JHEP01\(2019\)121](http://dx.doi.org/10.1007/JHEP01(2019)121) (backrefpages 97, 98).
- [123] P. W. Battaglia and others, *Relational inductive biases, deep learning, and graph networks*, CoRR (2018), URL: <http://arxiv.org/abs/1806.01261> (backrefpage 98).

- [124] S. Ioffe **and** C. Szegedy, *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*, CoRR (2015), URL: <http://arxiv.org/abs/1502.03167> (backrefpages 99, 196, 256).
- [125] F. Chollet **and** others, *Keras*, <https://keras.io>, 2015 (backrefpage 99).
- [126] M. Abadi **and** others, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, Software available from tensorflow.org, 2015, URL: <https://www.tensorflow.org/> (backrefpage 99).
- [127] B. Lakshminarayanan, A. Pritzel **and** C. Blundell, *Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles*, 2016, arXiv: 1612.01474 [stat.ML] (backrefpages 100, 162, 198, 201).
- [128] K. Simonyan, A. Vedaldi **and** A. Zisserman, *Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps*, 2013, arXiv: 1312.6034 [cs.CV] (backrefpage 101).
- [129] D. Koller **and** N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*, MIT press, 2009 page 51 (backrefpage 104).
- [130] ATLAS Collaboration, *Graph Neural Network Jet Flavour Tagging with the ATLAS Detector*, ATL-PHYS-PUB-2022-027, 2022, URL: <https://cds.cern.ch/record/2811135/> (backrefpage 106).
- [131] D. D. Wackerly, W. M. III **and** R. L. Scheaffer, *Mathematical Statistics with Applications*, 7th Edition, 2008 (backrefpage 110).
- [132] G. Cowan, K. Cranmer, E. Gross **and** O. Vitells, *Eur. Phys. J. C* **73** (2013) page 2501 (backrefpages 112, 113, 218, 219).
- [133] A. Wald, *Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large*, Transactions of the American Mathematical Society **54** (1943) page 426 (backrefpage 113).
- [134] A. L. Read, *Presentation of search results: the CL_S technique*, *J. Phys. G* **28** (2002) page 2693 (backrefpage 113).
- [135] ATLAS Collaboration, *Projected sensitivity of Higgs boson pair production combining the $b\bar{b}\gamma\gamma$ and $b\bar{b}\tau^+\tau^-$ final states with the ATLAS detector at the HL-LHC*, ATL-PHYS-PUB-2022-005, 2022, URL: <http://cdsweb.cern.ch/record/2799146> (backrefpage 117).
- [136] CMS Collaboration, *A portrait of the Higgs boson by the CMS experiment ten years after the discovery*, *Nature* **607** (2022) page 60, arXiv: 2207.00043 [hep-ex] (backrefpages 117, 187).
- [137] CERN, *CERN Yellow Reports: Monographs, Vol 2 (2017): Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector*, en, 2017, URL: <https://e-publishing.cern.ch/index.php/CYRM/issue/view/32> (backrefpage 117).

- [138] B. D. Micco, M. Gouzevitch, J. Mazzitelli and C. Vernieri, *Higgs boson potential at colliders: Status and perspectives*, *Reviews in Physics* **5** (2020) page 100045, URL: <https://doi.org/10.1016/j.revip.2020.100045> (backrefpages 117, 119).
- [139] M. Grazzini and others, *Higgs boson pair production at NNLO with top quark mass effects*, *Journal of High Energy Physics* **2018** (2018), ISSN: 1029-8479, URL: [http://dx.doi.org/10.1007/JHEP05\(2018\)059](http://dx.doi.org/10.1007/JHEP05(2018)059) (backrefpage 118).
- [140] F. A. Dreyer and A. Karlberg, *Vector-boson fusion Higgs pair production at N3LO*, *Physical Review D* **98** (2018), ISSN: 2470-0029, URL: <http://dx.doi.org/10.1103/PhysRevD.98.114016> (backrefpage 119).
- [141] I. Brivio and M. Trott, *The standard model as an effective field theory*, *Phys. Rept.* **793** (2019) page 1, arXiv: 1706.08945 [hep-ph] (backrefpage 125).
- [142] B. Grzadkowski, M. Iskrzyński, M. Misiak and J. Rosiek, *Dimension-six terms in the Standard Model Lagrangian*, *Journal of High Energy Physics* **2010** (2010), ISSN: 1029-8479, URL: [http://dx.doi.org/10.1007/JHEP10\(2010\)085](http://dx.doi.org/10.1007/JHEP10(2010)085) (backrefpage 126).
- [143] L. P. Sanchez, *ATLAS HH EFT Roadmap*, ATLAS HH Workshop, 2021 (backrefpage 126).
- [144] G. Buchalla, O. Catà and C. Krause, *Complete Electroweak Chiral Lagrangian with a Light Higgs at NLO*, *Nucl. Phys. B* **880** (2014) page 552, arXiv: 1307.5017 [hep-ph], Erratum: *Nucl. Phys. B* **913** (2016) 475 (backrefpage 126).
- [145] R. Alonso, M. Gavela, L. Merlo, S. Rigolin and J. Yepes, *The Effective Chiral Lagrangian for a Light Dynamical "Higgs Particle"*, *Phys. Lett. B* **722** (2012) page 330, arXiv: 1212.3305 [hep-ph] (backrefpage 126).
- [146] ATLAS Collaboration, *HEFT interpretations of Higgs boson pair searches in $b\bar{b}\gamma\gamma$ and $b\bar{b}\tau^+\tau^-$ final states and their combination in ATLAS*, ATL-PHYS-PUB-2022-019, 2022, URL: <https://cds.cern.ch/record/2806411/> (backrefpage 127).
- [147] A. Carvalho and others, *Higgs Pair Production: Choosing Benchmarks With Cluster Analysis*, *JHEP* **04** (2016) page 126, arXiv: 1507.02245 [hep-ph] (backrefpage 127).
- [148] M. Capozzi and G. Heinrich, *Exploring anomalous couplings in Higgs boson pair production through shape analysis*, *JHEP* **03** (2020) page 091, arXiv: 1908.08923 [hep-ph] (backrefpages 127, 128).
- [149] M. J. Dolan, C. Englert and M. Spannowsky, *Higgs self-coupling measurements at the LHC*, *Journal of High Energy Physics* **2012** (2012) (backrefpage 129).
- [150] ATLAS Collaboration, *Selection of jets produced in 13 TeV proton–proton collisions with the ATLAS detector*, ATLAS-CONF-2015-029, 2015, URL: <https://cds.cern.ch/record/2037702> (backrefpage 134).

- [151] J. Pearkes, *B-jet Energy Corrections and Improvements to the m_{bb} Resolution*, ATLAS HH Kick-Off Workshop, 2019 ([backrefpage 136](#)).
- [152] CMS Collaboration, *Search for narrow and broad dijet resonances in proton–proton collisions at $\sqrt{s} = 13$ TeV and constraints on dark matter mediators and other new particles*, [JHEP 08 \(2018\) page 130](#), arXiv: [1806.00843 \[hep-ex\]](#) ([backrefpage 142](#)).
- [153] ATLAS Collaboration, *Combination of searches for non-resonant and resonant Higgs boson pair production in the $b\bar{b}\gamma\gamma$, $b\bar{b}\tau^+\tau^-$ and $b\bar{b}b\bar{b}$ decay channels using pp collisions at $(\sqrt{s} = 13$ TeV with the ATLAS detector*, ATLAS-CONF-2021-052, 2021, URL: <https://cds.cern.ch/record/2786865> ([backrefpages 146, 147, 187](#)).
- [154] ATLAS Collaboration, *Search for the $HH \rightarrow b\bar{b}b\bar{b}$ process via vector-boson fusion production using proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, [JHEP 07 \(2020\) page 108](#), arXiv: [2001.05178 \[hep-ex\]](#), [Erratum: [JHEP 01 \(2021\) 145](#)] ([backrefpage 150](#)).
- [155] A. Ng, *Supervised learning: Classification and Logistic Regression*, Fall 2016 ([backrefpages 155, 242](#)).
- [156] ATLAS Collaboration, *Search for pair production of higgsinos in final states with at least three b-tagged jets in $\sqrt{s} = 13$ TeV pp collisions using the ATLAS detector*, [Phys. Rev. D 98 \(2018\) page 092002](#), arXiv: [1806.04030 \[hep-ex\]](#) ([backrefpage 155](#)).
- [157] G. V. Moustakides and K. Basioti, *Training Neural Networks for Likelihood / Density Ratio Estimation*, arXiv: [1911.00405](#) ([backrefpage 155](#)).
- [158] V. Nair and G. E. Hinton, *Rectified linear units improve restricted boltzmann machines*, 2010 ([backrefpage 155](#)).
- [159] E. Norrbin and T. Sjöstrand, *Production and hadronization of heavy quarks*, [The European Physical Journal C 17 \(2000\) page 137](#), URL: <https://doi.org/10.1007/s100520000460> ([backrefpage 155](#)).
- [160] Patrick Bryant, *Search for pair production of Higgs bosons in the 4 bottom quark final state using proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, CERN-THESIS-2018-208, 2018 ([backrefpage 156](#)).
- [161] S. Fort, H. Hu and B. Lakshminarayanan, *Deep Ensembles: A Loss Landscape Perspective*, 2019, URL: <https://arxiv.org/abs/1912.02757> ([backrefpage 162](#)).
- [162] J. Butterworth and others, *PDF4LHC recommendations for LHC Run II*, [J. Phys. G 43 \(2016\) page 023001](#), arXiv: [1510.03865 \[hep-ph\]](#) ([backrefpage 182](#)).
- [163] ATLAS Collaboration, *Constraining the Higgs boson self-coupling from single- and double-Higgs production with the ATLAS detector using pp collisions at $(\sqrt{s} = 13$ TeV*, ATLAS-CONF-2022-050, 2022, URL: <https://cds.cern.ch/record/2816332> ([backrefpages 187, 188, 234, 235](#)).

- [164] ATLAS Collaboration, *Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, [Phys. Rev. D **94** \(2016\) page 052002](#), arXiv: [1606.04782 \[hep-ex\]](#) (**backrefpage 188**).
- [165] ATLAS Collaboration, *Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, ATLAS-CONF-2016-049, 2016, URL: <https://cds.cern.ch/record/2206131> (**backrefpage 188**).
- [166] F. Pedregosa **and others**, *Scikit-learn: Machine Learning in Python*, Journal of Machine Learning Research **12** (2011) **page 2825** (**backrefpages 190, 193**).
- [167] C. E. Rasmussen **and** C. K. Williams, *Gaussian Processes for Machine Learning*, MIT press Cambridge, 2006 (**backrefpages 191, 192**).
- [168] C. B. Do **and** H. Lee, *Gaussian processes*, Fall 2016 (**backrefpage 191**).
- [169] *Inverse transform sampling*, wikipedia, URL: https://en.wikipedia.org/wiki/Inverse_transform_sampling (**backrefpage 193**).
- [170] D. J. Rezende **and** S. Mohamed, *Variational Inference with Normalizing Flows*, 2015, URL: <https://arxiv.org/abs/1505.05770> (**backrefpage 194**).
- [171] G. Papamakarios, *Neural Density Estimation and Likelihood-free Inference*, 2019, URL: <https://arxiv.org/abs/1910.13233> (**backrefpage 195**).
- [172] G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed **and** B. Lakshminarayanan, *Normalizing Flows for Probabilistic Modeling and Inference*, (2019), URL: <https://arxiv.org/abs/1912.02762> (**backrefpage 195**).
- [173] C. Durkan, A. Bekasov, I. Murray **and** G. Papamakarios, *Neural Spline Flows*, 2019, URL: <https://arxiv.org/abs/1906.04032> (**backrefpages 195, 196, 263**).
- [174] C. Durkan, A. Bekasov, I. Murray **and** G. Papamakarios, *nflows: normalizing flows in PyTorch*, version v0.14, 2020, URL: <https://doi.org/10.5281/zenodo.4296287> (**backrefpage 195**).
- [175] *SOFTPLUS*, PyTorch documentation, URL: <https://pytorch.org/docs/stable/generated/torch.nn.Softplus.html> (**backrefpage 196**).
- [176] K. He, X. Zhang, S. Ren **and** J. Sun, *Deep Residual Learning for Image Recognition*, 2015, URL: <https://arxiv.org/abs/1512.03385> (**backrefpage 198**).
- [177] L. Cadamuro, *Experimental Summary and Outlook*, Higgs Pairs Workshop, 2022 (**backrefpage 207**).
- [178] T. Kuhn, *The structure of scientific revolutions*, University of Chicago Press: Chicago, 1962 (**backrefpage 209**).
- [179] T. S. Collaboration, *A Measurement of R_b Using a Vertex Mass Tag*, [Physical Review Letters **80** \(1998\) page 660](#), URL: <https://doi.org/10.1103/physrevlett.80.660> (**backrefpage 213**).

- [180] Tom Wright, *Parity Violation in Decays of Z Bosons into Heavy Quarks at SLD*, SLAC-R-602, 2002 (**backrefpage 213**).
- [181] R. Barlow, *Practical Statistics for Particle Physics Lecture 3: Setting limits, claiming discoveries*, CERN European School on High Energy Physics, 2019 (**backrefpage 220**).
- [182] ATLAS Collaboration, *Search for Higgs boson pair production in the two bottom quarks plus two photons final state in pp collisions at $(\sqrt{s} = 13 \text{ TeV})$ with the ATLAS detector.*, ATLAS-CONF-2021-016, 2021, URL: <https://cds.cern.ch/record/2759683> (**backrefpage 234**).
- [183] ATLAS Collaboration, *Search for resonant and non-resonant Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ decay channel using (13 TeV) pp collision data from the ATLAS detector.*, ATLAS-CONF-2021-030, 2021, URL: <https://cds.cern.ch/record/2777236> (**backrefpage 234**).
- [184] Todd Seiss, *A search for the production of pairs Higgs bosons decaying to four bottom quarks with the ATLAS detector using $\sqrt{s} = 13 \text{ TeV}$ proton-proton collisions at the Large Hadron Collider*, CERN-THESIS-2021-049, 2018 (**backrefpage 236**).
- [185] Sean Gasiorowski, *Searches for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using the ATLAS detector, or: How I Learned to Stop Worrying and Love the QCD Background*, 2021 (**backrefpage 243**).
- [186] Rafael Teixeira de Lima, *How to Obtain a Pure QCD Reweighting*, Personal Communication, 2021 (**backrefpage 255**).
- [187] L. Dinh, J. Sohl-Dickstein and S. Bengio, *Density estimation using Real NVP*, 2016, URL: <https://arxiv.org/abs/1605.08803> (**backrefpages 260, 261**).
- [188] P. Kirichenko, P. Izmailov and A. G. Wilson, *Why Normalizing Flows Fail to Detect Out-of-Distribution Data*, 2020, URL: <https://arxiv.org/abs/2006.08545> (**backrefpage 264**).