Thesis Proposal for Stanford University
Katie Malone
23 August, 2012

My thesis in experimental high-energy physics will be performed as part of the SLAC ATLAS group, based partially at CERN, under the supervision of Su Dong. The thesis will document the search for one or more electrically neutral supersymmetric (SUSY) Higgs bosons produced in association with a bottom quark and decaying to a bottom-antibottom quark pair. This channel is an important one for determining whether the newly-discovered Higgs-like particle is supersymmetric, but holds significant challenges in understanding the background, calibration, and other features of a signature with several bottom quarks ("b" hereafter will refer to either a bottom or antibottom quark).

The theoretical motivation and general strategy for such an analysis are well-studied. In a Minimally Supersymmetric Standard Model (MSSM) scenario, there are five different Higgs bosons: the electrically neutral h^0, H^0 and A^0, and the electrically charged H+ and H-. This analysis searches for an electrically neutral particle decaying to 2 b quarks, but for reasons detailed below, it is only sensitive to particles above 200 GeV or so, which (if the 126 GeV particle discovered on 4 July 2012 is the h^0) makes this a search for H^0 and A^0. In a scenario where the SUSY parameter $\tan\beta$ is large, H^0 and A^0 do not couple strongly to bosons and accordingly they decay primarily to a pair of b quarks, with a branching ratio around 80% for a wide range of H^0/ A^0 masses (the other 20% go almost entirely to τ leptons) . Additional effects of large $\tan\beta$ are an enhanced cross section for bH^0/ bA^0, making it more likely to show up in a search, and the possibility of mass degeneracy between H^0 and A^0, effectively doubling the cross section. At the same time, the production cross section for the Standard Model (SM) Higgs being produced in association with a b quark is very small, so any signal seen in this channel would point toward a non-SM explanation (requiring the presence of the third b is also crucial to control the QCD background, as detailed more below). It follows that the three-b channel is a useful search channel for H^0/ A^0 and the observation (or lack thereof) of a particle in such a channel could be combined with information from other channels to indicate whether the presumed 126 GeV Higgs boson has supersymmetric siblings.

This analysis involves finding events with at least three b-jets and reconstructing the mass of the two jets which are the most likely to be the decay products of H^0/ A^0 (usually the leading two jets). A signal in such a channel would appear as a broad peak in the reconstructed mass spectrum of the selected b jets above a falling combinatorial background; the peak can then be parameterized to determine the statistical significance of the observation. Alternatively, if no peak is seen, one can use the limit on the production cross section to rule out some region in $\tan\beta$/mass(A^0) MSSM phase space.

This analysis will have several major components. One of the first steps is to model background and signal distributions using Monte Carlo (MC) simulations, and validating this MC. Validation includes looking at variable distributions that were created with different generators; for example, comparing Pythia and Sherpa signal MC to assess differences based on the initial state assumptions and calculation methods. Next, I will apply cuts on jet transverse momentum, jet position in the detector, jet quality, and the likelihood that a jet came from a b quark. For W, Z and top backgrounds, the cuts reduce the background by typically 95-99%. The remaining QCD background, though, is substantial (>90% of the overall background after cuts) and not well modeled by the MC, which motivates a data-driven QCD background estimation that must be implemented and validated. For a first pass of the analysis, to be completed during approximately the 2012-2013 school year, I plan to construct a likelihood function that takes several well-modeled kinematic quantities and finds which permutation of 2 b-jets in the data event (of the leading 3 b-jets) gives the most signal-like event shape, where I use signal MC to help me define what a "signal-like" event looks like. Then afterwards, I will apply a correction based on the MC distribution shapes that extrapolates from a control region with 2 b-jets to the signal region (which has 3 b-jets), to account for cases where I pick the wrong pair of signal b-jets or where I have a QCD background event that has 3 real b-jets in it. To make this partially data-driven method as effective as possible, I am performing studies now to find and/or construct variables that are both well-modeled in the MC (the MC distribution agrees with the data distribution) and that provide good discrimination (MC shows different shapes between signal and background events).

For a second pass of the analysis, to be completed in the second half of 2013 and to be the basis of my thesis, I plan to perform a more sophisticated analysis that studies in detail the rate at which light flavor and charm quark jets are misidentified as b-jets, and where in the event (i.e. the first, second, or third jet) these misidentified jets are most often found. Then I plan to construct a series of templates for the various significant QCD background samples, and apply a global fit to determine the relative proportions of the different templates. Last, I can double-check the template results by using b-jet-specific variables such as the mass and significance of the secondary vertex to apply any necessary corrections. This second pass of the analysis will be more intricate, and requires a finer understanding of how QCD background can fake a signal, but the payoff for the extra work will potentially be more sensitivity to a signal.

There will be a number of other effects to understand as well. In data, I have already studied many of the effects of the trigger used to select signal events (there are different triggers in 2011 and 2012, to reflect different conditions in the detector, but all triggers are based on the presence of multiple high-transverse-momentum b-jets), although more trigger studies will likely be needed to achieve sensitivity to masses below 300 GeV or so, where the cross sections are larger but the trigger thresholds start to cut into the signal. At the same time, for H^0/ A^0 masses above 200 GeV or so, the signal peak can show up as a very broad distribution on top of the exponentially decaying background. At a typical energy resolution for b-jets, which is around 15-20%, this means that the signal peak can have a width up to about 50 GeV, which makes the signal difficult to distinguish from just a mis-measurement of the background normalization. I am currently working on a project to improve the b-jet angular resolution, specifically by creating and applying corrections to the measured position of the jet, with the end goal of sharpening the resolution of a signal peak and making the analysis sensitive to smaller cross sections. Another important effect to understand is how often, and in what way, light/charm QCD jets can be mistagged as b-jets (and vice versa), and to help answer that question, I plan to join b-tagging experts in the fall of 2013 in an effort to study b-tagging in events with many tags.

These issues and others will be quantified as systematic uncertainties as I set a cross section limit or quantify a peak in the mass spectrum. The analysis will be developed using a "blind" technique, where I do not look at the full data set until all cuts have been finalized, all corrections applied, etc. so that I avoid bias as much as possible. In practice, this means studying sideband distributions as mentioned above, or only looking at a subset (~10%) of the data while developing the analysis (the exact choice of blinding method will be determined after consulting with ATLAS Higgs coordinators). Once the full dataset is unblinded, I can compute a 95% confidence limit for the bH^0/ bA^0 to bbb cross section, which can then also be interpreted as an exclusion of some region in the mass(A^0)-tanβ plane. Alternatively, and more excitingly, upon unblinding I could see a peak in the reconstructed mass spectrum, which would be one of the first (perhaps the first) indication of MSSM physics, and which would be a topic for many further studies into non-Standard Model physics as we use the Higgs system as a springboard for understanding the exciting new physics landscape that the MSSM would present.

I am currently based at CERN, and I have been here for about 16 months. I anticipate moving back to SLAC sometime in the spring or early summer of 2013, by which point the LHC will be turned off for about two of years for repairs and upgrades. The first pass of the Higgs analysis is aiming to have a preliminary result ready for the winter conferences in 2013 (which are held in February-March, but the approval for which starts in early January), with the second pass of more careful and polished studies being completed in late 2013 or early 2014 for presentation at conferences and publication. I plan to write up my thesis after I move back to SLAC, probably as I move the second part of the analysis through the ATLAS documentation and review process, so that I will be on track to graduate in 2014 or early 2015.