High Energy Physics Research at the CMS Experiment

Salvatore Rappoccio, PhD.

September 19, 2013

1 Introduction

The research proposed here will focus on finding solutions to the hierarchy problem of particle physics using novel jet reconstruction techniques to reconstruct highly-boosted standard model particles such as top quarks, W/Z and Higgs bosons. This research will be critical in the next phase of Large Hadron Collider (LHC) operations to maximize the discovery potential at the Energy Frontier of particle physics.

With the discovery of a standard model (SM)-like Higgs boson at 125 GeV during the 7 and 8 TeV runs of the LHC ("Run 1") [1, 2], the focus of collider physics now turns to understanding the nature of the observed mechanism for electroweak symmetry breaking. The SM is an effective theory, and quadratically-divergent quantum-loop corrections to the Higgs mass would require an enormous degree of fine tuning, if the Higgs mass is to remain finite up to the Planck scale. This is commonly referred to as the hierarchy problem, which can be resolved by introducing contributions beyond the standard model (BSM). These BSM models must survive a myriad of precision tests of the SM (such as the recent observation of $B_s \to \mu^+\mu^-$ [3, 4]), and extensive direct searches.

Various mechanisms beyond the standard model have been proposed to resolve the hierarchy problem. Since the most divergent quantum corrections to the Higgs mass involve top quarks and SM bosons, it is natural to suppose that BSM mechanisms preferentially involve loop interactions with these particles to cancel this divergence. Two major classes of models which solve the hierarchy problem discussed above are supersymmetry (SUSY) [5, 6, 7, 8, 9] and extra dimensions (ED) [10, 11, 12, 13]. After Run 1 of the LHC, the available phase space for these models contains SM particles in the final state that are often highly Lorentz boosted.

In recent years, new jet-clustering techniques have been developed to handle highly-boosted massive SM particles [14, 15, 16, 17, 18, 19, 20, 21, 22]. Summaries of these techniques are in Refs. [23, 24]. These large jets are referred to as "boosted jets" ("boosted tops", "boosted W's", "boosted Z's", "boosted Higgs"), and they are different from jets that originate from quantum chromodynamic (QCD) processes in that they exhibit different internal structure ("substructure"), and have an intrinsic jet mass. These algorithms analyze the substructure and mass of the boosted jet to break it up into smaller "subjets" which can be analyzed at smaller angular scales, thereby allowing analysis of highly-boosted states. Figure 1 shows an event display of such a highly-boosted top quark at CMS. The blue rectangles represent the energy measured in the hadronic calorimeter, the green rectangles represent the energy measured in the electromagnetic calorimeter, and the three groups of colored lines (purple, red, and orange) represent three different subjets as measured in the tracker. Using traditional reconstruction techniques, this object would be reconstructed as a single four-vector, whereas the newer techniques involving jet substructure are able to discern the three separate subjets.

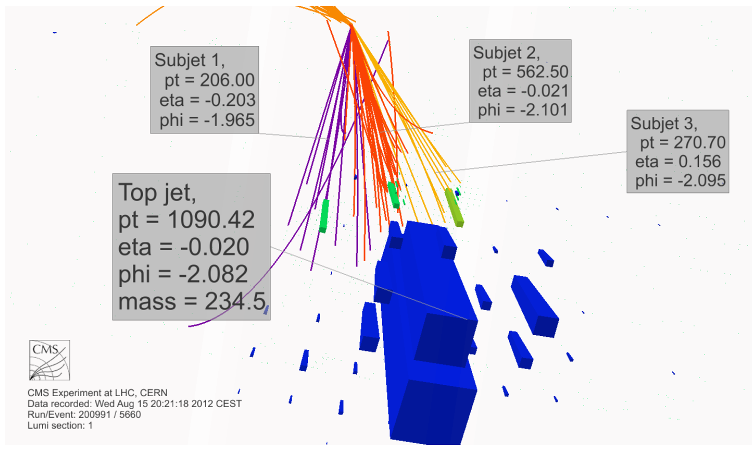

Figure 1: Event display of a highly-boosted top jet. The blue rectangles represent the energy measured in the hadronic calorimeter, the green rectangles represent the energy measured in the electromagnetic calorimeter, and the three groups of colored lines (purple, red, and orange) represent three different "subjets" as measured in the tracker.

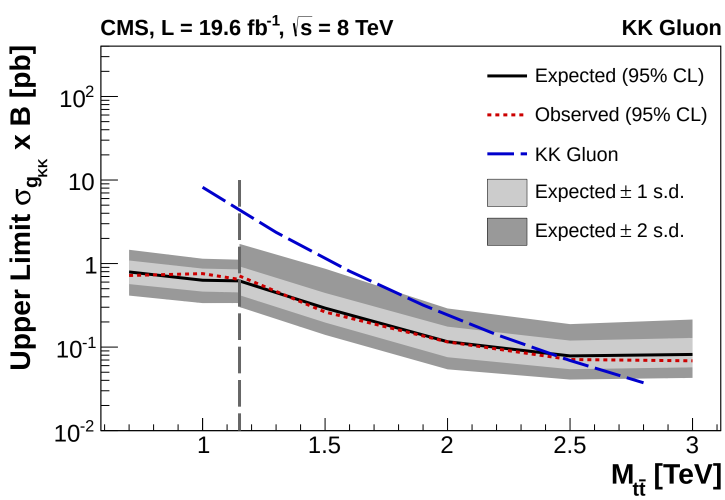

Figure 2: Limits on the mass of possible Kaluza-Klein excitations of the gluon in decays to $t\bar{t}$ pairs from a combination of results in Refs. [25, 26].

Without using these boosted-top strategies, the sensitivity of searches for particles predicted by these BSM models is reduced by a factor of ten or more, even at energies accessible during Run 1 of the LHC. For instance, Fig. 2 shows the cross-section limits for a KK gluon from the analysis in Ref [27]. At a mass of 2 TeV, the sensitivity of boosted-top techniques is a factor of 10 better than the sensitivity without using these techniques. This improvement grows even larger with the mass of the KK gluon.

2 Prior Work

2.1 Overview of Activities

Prof. Salvatore Rappoccio (tenure-track Assistant Professor) joined the Faculty at the University at Buffalo, SUNY (UB) in 2012. The UB group has an NSF grant (Award Number 1205960, July 15th 2012 - June 30th 2015) from Professors Kharchilava and Iashvili, however this was renewed one year before Prof. Rappoccio joined the Faculty and does not provide support for him or his group.

Prof. Rappoccio's group consists of Dr. James Dolen (postdoctoral fellow), Mr. Joshua Kaisen and Ms. Maral Alyari (graduate students), Mr. Brendan Smith and Mr. Jonathan Goodrum (undergraduate students, the latter supported under the "Collegiate Science and Technology Program" (CSTEP) at UB, which provides research funding for minority and disadvantaged undergraduate students). Work has already started on the proposed research using short-term startup funds. These startup funds will run out by 30-June-2014. If no funding is found, the support for Dr. Dolen, Mr. Kaisen, and Ms. Alyari will run out, and they will not be able to continue their studies.

The larger UB group consists of Profs. Kharchilava and Iashvili, two postdoctoral fellows, Dr. Ashish Kumar and Dr. Supriya Jain, and three Ph.D. students, Mr. Joseph Zennamo, Mr. Andrew Godshalk, and Mr. Jimin George. Prof. Rappoccio's group has already commenced collaboration with the teams of Kharchilava and Iashvili, specifically in the commissioning of the pixel detector.

2.2 CMS Activities

Prof. Rappoccio has played a critical leading role in implementing and understanding these jet substructure and boosted techniques at the LHC in general, and became the foremost expert of this topic at CMS. This work started while he was a postdoctoral fellow at Johns Hopkins University working under NSF Grant 1100862, and expanded greatly during his tenure at UB, based on his short-term startup funds.

Rappoccio's early studies (before collision data) and implementation of the algorithms are outlined in Ref. [17]. He has authored the first searches for new physics with collision data to be published using boosted top [28] and boosted W/Z [29] techniques anywhere in the world, as well as the first measurement of the jet mass at CMS [30]. In addition, he has contributed to several of the now-seminal review papers about the subject, which are used as standard references and benchmarks in the community at large [23, 24].

After these pioneering efforts successfully established the CMS Collaboration as a world leader in the fields of jet substructure and boosted hadronic final states, Prof. Rappoccio was made the convener of a newly-formed "Beyond Two Generations" Physics Analysis Group (B2G PAG) at CMS, which focuses on these highly-boosted final states, in particular in the top sector. In his tenure as convener since August 2012, CMS has deployed these techniques in 4 analyses of the 8 TeV data (one paper submitted [27], and several in preparation) [31, 32, 33]. Furthermore, the techniques are being adopted in four other analyses to search for other exotic signatures (including the 8-TeV continuation of Ref. [29] in Ref. [34]), as well as two analyses to search for supersymmetry.

For instance, the latest 8 TeV limits using boosted techniques to search for $t\bar{t}$ resonances of Kaluza-Klein gluons from Ref. [27] are shown in Fig. 2 and finally probe, for the first time, a viable region of parameter space available for models of ED that are not already disfavored by precision measurements [35]. Without the boosted-jet strategies for searches that Prof. Rappoccio has developed, it would not have been possible to access this regime at CMS by any other method. This highlights the absolutely critical nature of developing these tools in searches.

In addition to pioneering the searches for new physics using boosted-jet techniques, Prof. Rappoccio has also been instrumental in developing, commissioning, and maintaining the general jet reconstruction and particle-flow algorithms at CMS. Some of the areas in which he has played a leading role are the development of the "charged hadron subtraction" algorithm at CMS, where charged PF candidates associated with pileup vertices were removed directly from jets, resulting in a 10-20% improvement in the jet energy resolution, as well as the development of the median-p_T-per-unit-area pileup correction for jets at CMS. He was also instrumental in getting the "anti-k_T" algorithm [36] accepted by the CMS collaboration, and redesigned the software framework to make usage of the `fastjet` package [37, 38] so as to have a common algorithmic basis for jet reconstruction in ATLAS, CMS, and the theoretical community. He also participated in the software validation and data-quality monitoring (DQM) of the jets at CMS before and during the Run 1 startup commissioning at 2.36 and 7 TeV.

Furthermore, Prof. Rappoccio's group has participated in the larger UB activities of FPIX commissioning and testing during the first year of his position. Dr. Dolen and Dr. Kumar developed tests for the FPIX high-density interconnect (HDI) modules using a probe station at Fermilab, leading a team of graduate and undergraduate students after severe budgetary constraints at Fermilab significantly reduced the number of people working on this particular part of the project.

Another past responsibility of Prof. Rappoccio at CMS was the Level-2 management position of the "Analysis Tools" subgroup of the Offline Project. During his tenure, Run 1 of the LHC began, and Rappoccio was responsible for guiding a team of people to ensure that proper analysis of the collision data could be done. This included such activities as minimizing the size of the "Analysis Object Dataset" (AOD) data tier, the primary access point for all CMS analyses, the development of the "Physics Analysis Toolkit" [39], and development and maintenance of the `FWLite` software framework, which is a small library of the CMS software tools (`CMSSW`) designed for usage by analysts.

Finally, Prof. Rappoccio has been involved in the review of 11 separate analyses at CMS as a member of the "Analysis Review Committee" (ARC).

2.3 Non-CMS Activities

To highlight the importance of the proposed research, the "SnowMass on the Mississippi" Community Summer Study [40] effort has made the identification of boosted jets an extremely high priority for the future of HEP in general. Dr. Dolen has been extensively involved in this effort, and is co-leading the group that focuses on boosted top-quark tagging techniques. The Snowmass "white paper" describing the need for these techniques, and the capability of future detectors to utilize them, is described in Ref. [41].

In addition, Prof. Rappoccio has played an integral part of the **BOOST** conference series and other conferences, such as the **Boston Jet Substructure Workshop**. The BOOST conference is the foremost conference on jet substructure and boosted-jet algorithms, and the conference reports in 2010 and 2011 are used as standard references in benchmarking these techniques [23, 24].

As a graduate student at Harvard University, Rappoccio also performed major updates and modifications of the Silicon Readout Controller (SRC) for the CDF experiment at Fermilab. These developments allowed the readout of the silicon detector at Level 1 of the CDF triggering system, which then could be passed to a tracking algorithm at Level 2. This made the extensive B-physics program at CDF possible, since without these track triggers, the quality of results would have been severely degraded.

Finally, for his expertise in the field of jet substructure, Rappoccio was the primary reviewer for the ATLAS paper describing jet substructure and boosted techniques in Ref [42], submitted to the *Journal of High Energy Physics* (JHEP).

3 Proposed Program of Research

The proposed program of research outlined here will have three separate timelines. The first is to finish the analysis of the 8 TeV collision data from Run 1 of the LHC. The second is to ensure that the discovery capacity of CMS using boosted jets is maximized at the onset of the 13 TeV run of the LHC ("Run 2"). The third is to further develop the capabilities of CMS at all levels (from detectors to analysis of data) in long-term planning of the high-luminosity LHC (HL-LHC, or "Run 3").

The focal point of this proposal is to pursue searches for BSM physics with boosted objects, described in Sec. 3.1.1. In addition to BSM searches, it is also important to characterize the SM contributions in these samples, and a program of study is described in Sec. 3.1.2.

For these goals to be met throughout the three LHC runs, many technical challenges must be met. The development and deployment of the particle flow [43] algorithm is discussed in Sec. 3.2.2. The collection and monitoring of data collection are outlined in Sec. 3.2.3. Finally, longer-term detector upgrades to the forward pixel (FPIX) detector are discussed in Sec. 3.2.4.

3.1 Proposed Physics Analyses

3.1.1 BSM Searches with Boosted Jets

The major motivation of this research proposal is the discovery of BSM physics with boosted jets. This is an ambitious goal, but one that Prof. Rappoccio is extraordinarily well-equipped to handle.

Rappoccio's tenure as "Beyond Two Generations" (B2G) Physics Analysis Group Convener at CMS continues until the end of 2014. The goals during this initial stage of the proposed research are to finish the analyses of the 8 TeV CMS data searching for BSM physics, to prepare in earnest for the upcoming LHC Run 2, and to motivate longer-term upgrades for LHC Run 3 ("Phase 1" and "Phase 2" CMS detector upgrades).

The immediate goal is to finish the analysis of well-motivated signals in the B2G group. The most important of these include searches for $\mathrm{t\bar{t}}$ resonances and heavy-quark partners. These are two clear signatures of ED that can be observed at CMS. The research program of the B2G group, under Prof. Rappoccio's guidance, includes several seminal "legacy" papers on these topics, to offer a final word on the 8 TeV data at the LHC. In addition, several other signatures are being investigated, including long-lived top quarks, dark matter signatures involving top quarks, and baryon-number violation in top quark decays.

There are two major analyses that Prof. Rappoccio's group has especially focused on in the past, and will continue to focus on in the future. Prof. Rappoccio's group are leading analysts on the $\mathrm{t\bar{t}}$ BSM searches, including the latest submission of a search in this channel [27]. While the current limits on BSM production of $\mathrm{t\bar{t}}$ are stringent, further improvements are possible with the Run 1 data. The full suite of existing jet substructure tools that Prof. Rappoccio has developed over the years (as described in Sec. 2.2), along with new additional techniques involving the combination of jet substructure and bottom-quark jet tagging, are expected to yield significant improvements over expected sensitivities in the $\mathrm{t\bar{t}}$ resonance search, of up to 30%.

Additionally, Prof. Rappoccio's group will continue to contribute to searches for diboson resonances ($WW/WZ/ZZ$) in the CMS Exotica group. Many advanced jet reconstruction techniques have already been deployed in this channel, but the paper of the combination of all of the various channels has not yet been completed. The research proposed here will help to bring this sensitive analysis to fruition.

In addition to the completion of the analysis of the LHC Run 1 data, the preparations for the startup of the LHC for Run 2 must be run concurrently. There are a suite of technical challenges to address in order for this to be accomplished, described in detail below in Sec. 3.2.

3.1.2 SM Measurements with Boosted Jets

The searches performed during the first run of the LHC have not yielded signals of BSM physics. However, these data can still be used to perform measurements of SM particles.

In particular, the differential cross section for $\mathrm{t\bar{t}}$ production at CMS [44] is sensitive to the case where top quarks are produced with resolved decay products. However, this limits the sensitivity of the measurement to top quarks with low p_{T}. The boosted-top identification techniques outlined in Sec. 3.1.1 can also select a very pure sample of $t\bar{t}$ pairs produced via SM mechanisms, and hence can also provide an opportunity to measure $t\bar{t}$ production in the highly-boosted regime where the traditional techniques have no sensitivity.

The measurement of the $t\bar{t}$ cross section in this regime is very interesting for a number of reasons. Firstly, it provides a sensitive test of perturbative QCD predictions for the NNLO $t\bar{t}$ production cross section that is not tested elsewhere. In addition, it may provide additional constraints on parton distribution functions (PDFs) for high-x partons. Finally, since the production of $t\bar{t}$ pairs at this kinematic range is dominated by $q\bar{q}$ production, tests of the forward-backward asymmetry could be measured to compare to those measured at the Tevatron [45]. The existing LHC measurements focus on the central-forward charge asymmetry [46, 47] because they are performed at low top-quark p_{T}, and are not sensitive to the same physics quantity. Explorations of this regime could provide an interesting point of comparison and may elucidate the anomalously large forward-backward asymmetry observed at the Tevatron.

Prof. Rappoccio is already leading a large effort at CMS to measure the $t\bar{t}$ production cross section for $p_{\mathrm{T}}^{\mathrm{top}} > 400$ GeV. This effort grew out of a CMS educational program, the CMS Data Analysis School at Fermilab in January 2013. Under Rappoccio's direction, a team of three faculty, one postdoc, four graduate students and four undergraduate students from UB, Cornell, Siena College, and Johns Hopkins are producing a measurement of the $t\bar{t}$ production cross section, fully corrected for detector effects.

3.2 Technical Challenges

3.2.1 Boosted Jet Algorithms

The performance of the boosted-jet algorithms at CMS will be heavily affected by increasing pileup in Run 2 and beyond. For instance, Fig. 3 was made by Dr. Dolen in the context of Snowmass (as described above), and shows the jet mass of a boosted top-quark jet in three separate cases during Run 3: without pileup (solid histogram), with 140 pileup interactions (open green histogram), and with 140 pileup interactions after applying an advanced boosted jet "grooming" technique (the "jet trimming" technique [22], in the open magenta histogram). The naive implementation of the "Run 1"-style algorithm to a higher pileup regime fails miserably. The major innovation of combining strengths from different techniques, however, can rescue the performance quite handily, and restores the jet mass almost back to the case where there was no pileup at all.

This is one example of a problem that was already identified and fixed by using innovative techniques to resolve the existing problems. There are several other major limitations that are also under investigation which will also require a heavy amount of development to solve for Run 2 and beyond. For example, boosted-jet techniques start to break down at very high energies because the SM particles are sufficiently boosted that their decay products fall into one single calorimeter cell. This could be mitigated, for instance, with a more finely-tuned PF algorithm as described in the next section. Only through a robust and comprehensive development and deployment strategy will these challenges be met.

The work outlined in this proposal includes the development of new techniques in order to fully realize the capability of the CMS detector in discovering new physics with boosted jets in an environment that is increasingly difficult to analyze. The success of this approach will require major innovations and cutting-edge tools, as well as a broad view of the field at large. The previous experience of Prof. Rappoccio has provided the expertise needed to solve these extensive problems, and can ensure that timely and deployable strategies can be developed prior to the startup of Run 2 of the LHC.

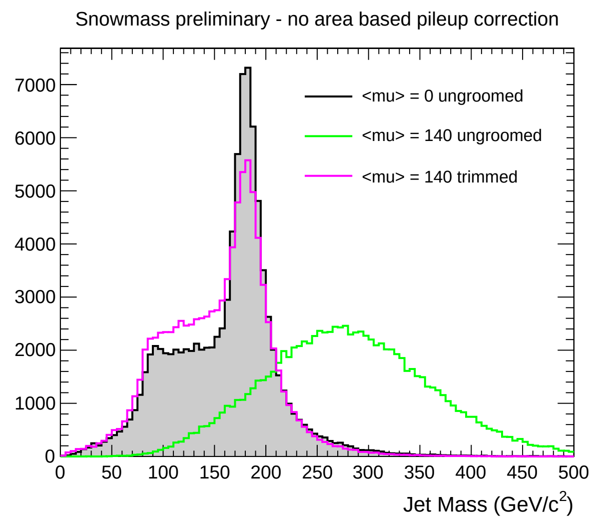

Figure 3: Jet mass for boosted top-quark jets produced at $\sqrt{s} = 8$ TeV. Three cases are shown : without pileup (solid histogram), with 140 pileup interactions (open green histogram), and with 140 pileup interactions after applying the jet trimming technique (open magenta histogram).

3.2.2 Particle Flow

The success of the general physics program at CMS during the first run of the LHC, and in particular of the jet substructure techniques outlined above, were heavily contingent on the PF algorithm [43]. When using the PF algorithm, performance of the top-tagging algorithms at CMS was increased by 20% compared to using calorimetric information only, driven by the superior energy and angular resolution of the algorithm. In addition, the PF algorithm is capable of removing over 60% of pileup directly, by removing charged hadrons that are associated to pileup vertices. This already reduces the scale of the problem of pileup considerably, and the substructure algorithms at CMS are heavily reliant on this technique.

The PF algorithm, however, is heavily tuned (by construction) on the Run 1 detector. A considerable amount of work must be done to migrate the PF algorithm to the new detector conditions, as well as to adapt to the enormous increase in pileup. In the coming “Phase 1” and “Phase 2” tracker and HCAL upgrades at CMS [48, 49] the detector will be improved in a number of ways, including the addition of a fourth silicon pixel layer to improve pileup rejection, and updates to the HCAL electronics to allow for readout of the longitudinal segmentation of hadronic showers. These upgrades will necessitate another tuning of the PF algorithm, with new opportunities and challenges directly ahead. Mr. Kaisen and Prof. Rappoccio are already working with the "global event description" (GED), tracking and HCAL groups at CMS to update the PF algorithm to take advantage of the new segmented HCAL readout, as well as the new pixel-detector configurations.

The success of the CMS boosted-object program is heavily contingent on the success of the PF algorithm. Without the superior energy and angular resolution of subjets and the pileup reduction capabilities of the PF algorithm, the substructure techniques we have developed will be significantly degraded.

However, with the newer detector comes new possibilities. It may be possible to mitigate the efficiency loss of jet substructure algorithms at very high energies (when the Lorentz boost is so large that all of the products decay in a single calorimeter cell) using the upgraded detector capabilities. For instance, combining information from the longitudinally-segmented readout of the HCAL with the finer angular segmentation of the ECAL and the trackers, it may be possible to lessen the impact of the intrinsic HCAL granularity. The upgraded CMS detector needs to be fully-utilized in order to maximize the potential for physics discoveries. Advanced development of the PF algorithm may illuminate new advantages that were previously impossible.

3.2.3 Trigger Development and Data Quality Monitoring

In order to analyze highly-boosted topologies in BSM searches, the relevant data must be collected in efficient triggers. Furthermore, the reconstruction, calibration, and trigger efficiencies of these algorithms must be incorporated in the overall program of data-quality monitoring (DQM) at CMS. At the present time, there are very few triggers that are capable of collecting boosted-jet events in the upcoming Run 2, and those that do exist necessitate very high kinematic thresholds which limit their sensitivity. This needs immediate rectification if we are to be prepared to adequately perform these searches for new physics with boosted objects detailed in this proposal.

A major difficulty here is that the jets in events in Run 2 will be highly-polluted with pileup interactions. In fact, using algorithms that we have previously developed in Run 1 to deal with pileup mitigation may not continue to work in Runs 2 and 3 with such large pileup. Since there is a large stochastic fluctuation of the pileup background (roughly 1 GeV per jet), with 140 pileup interactions, a 400 GeV jet will be expected to be composed of nearly 50% pileup.

Rigorous and expansive application of boosted-jet, jet-substructure, pileup removal, and other advanced techniques will be necessary to collect the boosted-jet datasets with reasonable efficiency without having enormously high event rates. For instance, if one were to select events with a jet-mass requirement, without any corrections the trigger rate would grow linearly with the number of pileup interactions, and without any corrections the trigger would rapidly become untenable. However, from the studies by Prof. Rappoccio in Ref. [30], it is observed that applications of grooming techniques to the jets can eliminate the growth of the jet mass with increasing pileup entirely. Further investigations in this avenue of inquiry are sure to yield extremely promising results.

Dr. Dolen and Ms. Alyari will partake in the analysis and development of the triggering program at CMS to ensure that boosted-jet searches will acquire the necessary data in Run 2 without having event rates that simply scale linearly with pileup. Furthermore, Ms. Alyari will be participating in the DQM of generic and boosted jets to ensure that the algorithms are being implemented effectively and that no adverse interactions occur unexpectedly, thereby ensuring the safety of the physics program proposed here.

3.2.4 Forward Pixel Detector Commissioning and Operations

In order to maintain the ability of the PF algorithm to perform pileup rejection in the foreseen high-luminosity run of the LHC (HL-LHC), good vertexing capabilities of the CMS detector are of the utmost importance. In addition to providing secondary-vertex identification in heavy-flavor jets, the upgraded pixel detector at CMS provides better efficiency for pileup track rejection, which is relied upon heavily in the aforementioned physics program. The planned "Phase 1" upgrade of the forward pixel (FPIX) detector at CMS is expected to add an additional pixel layer to improve fake track rejection, as well as to move the innermost pixel layer closer to the beam to provide a more sensitive impact parameter measurement.

Rappoccio's group, as well as the larger UB group and others in the US CMS program, are very active in the forward pixel (FPIX) detector commissioning for the Phase-1 upgrade. The UB group is involved with the US CMS NSF Cooperative Agreement "U.S. CMS Phase-1 Upgrades", Fastlane number 7334626, to support undergraduate students working on the CMS Phase 1 FPIX upgrade project. However, this Cooperative Agreement does not provide support for the graduate students nor postdoctoral fellows already working on the project.

In the past year, Rappoccio's group has begun working on the FPIX upgrade project. Dr. Dolen is permanently stationed at Fermilab to take advantage of the Silicon Detector (SiDet) Facility, and is already participating in the FPIX module testing. During the summer of 2013, the UB team (led by Dr. Dolen and Dr. Kumar) commissioned 50 High-Density Interconnect (HDI) boards for the FPIX modules, including visual inspection, spot checks of the connectivity of anomalies found by this visual inspection, and also randomized spot checks of interconnections to ensure high quality of the produced modules. For this exercise, a probe station was revived for usage at SiDet, and a LabView program was developed to do the testing.

In addition to the already completed activities, Rappoccio's group has committed to other HDI burn-in tests, tests of the half-disk assembly of the FPIX detector, beam tests (the first scheduled in Fall 2013), and overall commissioning of the FPIX detector. Longer-term activities will include operations and support of the upgraded Phase-1 pixel detector once it is installed at CMS.

4 Outreach and education

While it is critical to pursue a rigorous research program, a large part of the responsibility of scientists is to educate the next generation effectively. There is already extensive work being done to educate high school-level students and teachers via the *QuarkNet* program at UB, however there is very little in the way of educating the general public. In addition to participating in the existing *QuarkNet* activities, the plan outlined in this proposal will extend the coverage of the outreach program at UB to engage the broader public in discussions of major results in particle physics, as well as to enliven particle physics for young students. This will be implemented based on similar events as the "HiggsFest" [50] that Prof. Rappoccio organized at UB.

4.1 Higgsfest and other public events

The "Higgsfest" that was organized here at UB in 2012 is highlighted in Ref. [50]. The aim of the event was to invite the general public for "plain English" summaries and hands-on demonstrations that were geared for a multitude of age and knowledge levels. This was attended by over 100 people, including children, high-school students, physics and non-physics undergraduates, and interested members of the community.

Some of the hands-on demonstrations included building models of Feynman diagrams from craft material (for young children), a fully-functional four-layer coincidental muon scintillator detector, a cloud chamber made out of tupperware, felt, and dry ice, and the actual Higgs events from the CMS collaboration in an interactive event display. The event was covered by the "UB Reporter" here at UB [51].

Two more such events are proposed, the first to coincide with the LHC turn-on sometime in 2015, and the second to coincide with the newest results from the LHC after data-taking commences. These are events that should generate high media coverage, and will be a good opportunity to capitalize on public interest in this field. Having regular events to discuss the LHC results is a very long-term goal, and the opportunity to develop them with the CAREER proposal will be very useful. In the event of a major new discovery at the LHC during Run 2, the public interest will be very high, so having the experience of what works and what does not work in such events is extremely valuable to maximize the public impact.

In addition, this removes the stigma associated with science and technology fields at an early age. When young children can attend an event with their parents and take something away from it, this shows them that science is an integral part of life, and nothing to be particularly nervous about pursuing. It may even convince younger people to pursue a scientific career.

One of the major points learned during the last "Higgsfest" is that it is often difficult to have economically-disadvantaged students attend the lectures because of a lack of transportation possibilities. This is something to rectify for future projects along these lines. Therefore, in addition to holding the event directly at the UB North Campus (which is difficult for inner-city Buffalo schools to reach), a duplicated event is also proposed closer to the inner city that is easier to attend, or possibly to visit these schools directly. Some possible locations are the UB South Campus, or at "Babeville" [52], where the UB Physics

Department routinely organizes the “Science and Art Cabaret” [53]. Both locations provide the infrastructure needed for the event, and access for disadvantaged schools and students in the inner city of Buffalo.

4.2 Undergraduate research

Having been an undergraduate researcher, Prof. Rappoccio knows the importance of making a positive impression on undergraduates who are interested in pursuing an academic career. He is currently advising two undergraduate students in his group, Brendan Smith and Jonathan Goodrum. Mr. Goodrum is participating in the “Collegiate Science and Technology Entry Program” (CSTEP) here at UB. This program focuses on disadvantaged or minority students who would otherwise have a difficult time pursuing STEM-related fields. Prof. Rappoccio is a strong believer in the principles and practice of this program, and intends to continue this work in the future. Currently, Mr. Goodrum is participating in a study to increase the processing speed of jet-clustering algorithms (critical for the jet substructure outlined above) by investigating possible parallelization of the algorithms for multicore, multithread, and possibly graphical processing unit usage.

In addition, Mr. Smith obtained a complete hands-on immersion at the Fermi National Accelerator Laboratory (FNAL) during the summer of 2013. There, he participated in testing of major electronics modules for the forward pixel (FPIX) upgrade at CMS. This kind of experience is invaluable in the field of young researchers, and it is also planned to continue supporting undergraduate students in such hands-on activities at CERN next year under the NSF Cooperative Agreement that is under review at the present time, discussed in the “Current and Pending Support” section below (Proposal Number 7334626). Mr. Smith is also currently completing an undergraduate honor’s thesis under the direction of Prof. Rappoccio.

5 Prior Work in Education and Outreach

As Prof. Rappoccio wholeheartedly believes in the importance of outreach and educational activity, he has extensively participated in outreach activities throughout his career. Some examples are

- **Facilitator, CMS Data Analysis School, Fermilab, Batavia IL (Jan 2013, Jan 2011)** : facilitated the education of new CMS members through tutorials and hands-on exercises of real-life measurements.
- **“Higgsfest” public lecture, Buffalo NY (Dec 6 2012)**: as described above, the “Higgsfest” was a celebration of the discovery of the SM-like Higgs boson at the LHC focused on outreach to the community.
- **Science Cabaret public lecture, Buffalo NY (Oct 17 2012)**: lecture about the statistics behind the discovery of the Higgs boson aimed at conveying the information to a group of members of the artistic and public communities in Buffalo.

- **Fermilab UEC Trip to Washington, D.C. (Spring 2009, Spring 2011)** : advocacy trip to visit members of Congress to spread information related to particle physics.

- **Angels and Demons public lecture, Bethlehem Public Library, Delmar NY (May 15 2009)** : related to the release of the film "Angels and Demons," a large-scale public lecture series was organized by the Fermilab outreach department, and Prof. Rappoccio gave the lecture at a library outside of Albany, NY.

- **Chicago Section of the Society for Applied Spectroscopy, Elmhurst, IL (Jan 13 2009)** : lecture of the methodologies and results of collider physics to an engineering society in the Chicago Area.

- **Lecture, Bethlehem Central High School, Delmar NY (Dec 4 2008)** : lecture to high-school students from the high school I graduated from about particle physics and the LHC.

- **Duckon Science Fiction Convention, Naperville IL (Jun 14 2008)** : lecture about the "real" science of the LHC compared to science fiction.

- **Starter Kit for LHC Newcomers and Physics Analysis Toolkit Tutorials, Fermilab, Batavia IL, also CERN, Geneva CH (throughout 2008-2009)** : as part of his duties for the LHC support team at Fermilab, Prof. Rappoccio developed an education program for students, postdoctoral fellows, and faculty that were new to the CMS experiment with hands-on examples, simple walk-through tutorials, and in-person support. This eventually developed into full-fledged activities at CERN for the tutorials, as well as was the precursor for the hugely successful "CMS Data Analysis School," as much of the material was initially developed there.

- Prof. Rappoccio currently oversees the research and education of **Joshua Kaisen** and **Maral Alyari** as graduate students in his group. Mr. Kaisen is working on upgrading the particle-flow algorithm to work in the upgraded CMS detector. **Ms. Alyari** will be working on data-quality monitoring of the particle-flow and jet-reconstruction algorithms.

- Previously, at JHU, Prof. Rappoccio also oversaw the graduate studies of **Guofan Hu, Marc Osherson, Kevin Nash, and Yongjie Xin** who are working on searches for BSM physics using boosted jet algorithms. Dr. Hu graduated recently and moved to a position in finance, while Mr. Osherson, Mr. Nash, and Mr. Xin are continuing their studies, still partially under his supervision.

- Prior to the monitoring of the activities of **Brendan Smith** and **Jonathan Goodrum**, at JHU Prof. Rappoccio monitored the activities of two undergraduate students, **David Bjergaard** and **Prateek Bajaj**. Mr. Bjergaard is now at Duke University pursuing a Ph.D. in particle physics.

6 Summary

In summary, the program that has been proposed here is ambitious and high-profile, but one in which Prof. Rappoccio has a unique talent set to accomplish. This program will have an excellent chance of uncovering BSM physics during Run 2 of the LHC and beyond. The previous experience of Prof. Rappoccio with boosted-jet techniques will be absolutely critical to ensure that the techniques are as successful in Run 2 as they were in Run 1, amidst a sea of technical challenges.

In addition to this ground-breaking research, an addition to the extensive outreach at UB is proposed in events such as successors to "Higgsfest" in the Buffalo area (including lower-income places that have limited access to suburban venues).

References

[1] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 2012.

[2] Serguei Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B*, 2012.

[3] Serguei Chatrchyan et al. Measurement of the Bs to mu mu branching fraction and search for B0 to mu mu with the CMS Experiment. 2013.

[4] R Aaij et al. Measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and search for $B^0 \to \mu^+\mu^-$ decays at the LHCb experiment. 2013.

[5] J.F. Gunion, H.E. Haber, and M. Sher. Charge / Color Breaking Minima and a-Parameter Bounds in Supersymmetric Models. *Nucl. Phys. B*, 306:1, 1988.

[6] Jonathan L. Feng, Konstantin T. Matchev, and Takeo Moroi. Multi - TeV scalars are natural in minimal supergravity. *Phys. Rev. Lett.*, 84:2322, 2000.

[7] Ryuichiro Kitano and Yasunori Nomura. Supersymmetry, naturalness, and signatures at the LHC. *Phys. Rev. D*, 73:095004, 2006.

[8] Riccardo Barbieri and Duccio Pappadopulo. S-particles at their naturalness limits. *JHEP*, 0910:061, 2009.

[9] D. Horton and G.G. Ross. Naturalness and Focus Points with Non-Universal Gaugino Masses. *Nucl. Phys. B*, 830:221, 2010.

[10] Nima Arkani-Hamed, Savas Dimopoulos, and G. R. Dvali. The hierarchy problem and new dimensions at a millimeter. *Phys. Lett. B*, 429:263–272, 1998.

[11] Lisa Randall and Raman Sundrum. A large mass hierarchy from a small extra dimension. *Phys. Rev. Lett.*, 83:3370–3373, 1999.

[12] Lisa Randall and Raman Sundrum. An alternative to compactification. *Phys. Rev. Lett.*, 83:4690–4693, 1999.

[13] Kaustubh Agashe, Alexander Belyaev, Tadas Krupovnickas, Gilad Perez, and Joseph Virzi. LHC signals from warped extra dimensions. *Phys. Rev. D*, 77:015003, 2008.

[14] Michael H. Seymour. Searches for new particles using cone and cluster jet algorithms: A Comparative study. *Z. Phys. C*, 62:127, 1994.

[15] Jonathan M. Butterworth, Adam R. Davison, Mathieu Rubin, and Gavin P. Salam. Jet substructure as a new Higgs search channel at the LHC. *Phys. Rev. Lett.*, 100:242001, 2008.

[16] David E. Kaplan, Keith Rehermann, Matthew D. Schwartz, and Brock Tweedie. Top Tagging: A Method for Identifying Boosted Hadronically Decaying Top Quarks. *Phys. Rev. Lett.*, 101:142001, 2008.

[17] CMS Collaboration. A cambridge-aachen (c-a) based jet algorithm for boosted top-jet tagging. CMS Physics Analysis Summary CMS-PAS-JME-009-01, 2009.

[18] Stephen D. Ellis, Christopher K. Vermilion, and Jonathan R. Walsh. Techniques for improved heavy particle searches with jet substructure. *Phys. Rev. D*, 80:051501, 2009.

[19] Stephen D. Ellis, Christopher K. Vermilion, and Jonathan R. Walsh. Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches. *Phys. Rev. D*, 81:094023, 2010.

[20] Jesse Thaler and Ken Van Tilburg. Maximizing Boosted Top Identification by Minimizing N-subjettiness. *JHEP*, 1202:093, 2012.

[21] Tilman Plehn, Michael Spannowsky, Michihisa Takeuchi, and Dirk Zerwas. Stop Reconstruction with Tagged Tops. *JHEP*, 1010:078, 2010.

[22] D. Krohn, J. Thaler, and L. .T. Wang. Jet Trimming. *JHEP*, 02:084, 2010.

[23] A. Abdesselam, E. Bergeaas Kuutmann, U. Bitenc, G. Brooijmans, J. Butterworth, et al. Boosted objects: A Probe of beyond the Standard Model physics. *Eur. Phys. J. C*, 71:1661, 2011.

[24] A. Altheimer, S. Arora, L. Asquith, G. Brooijmans, J. Butterworth, et al. Jet Substructure at the Tevatron and LHC: New results, new tools, new benchmarks. *J. Phys. G*, 39:063001, 2012.

[25] Serguei Chatrchyan et al. Search for Anomalous Top Quark Pair Production in the Boosted All-Hadronic Final State using pp Collisions at sqrt(s) = 8 TeV. CMS-PAS-B2G-12-005 .

[26] CMS Collaboration. Search for $\mathrm{t\bar{t}}$ resonances in semileptonic final state. *CMS Physics Analysis Summary*, CMS-PAS-B2G-12-006, 2012.

[27] Serguei Chatrchyan et al. Searches for anomalous $t\bar{t}$ production in pp collisions at $\sqrt{s}$ =8 TeV. 2013. Submitted to PRL.

[28] Serguei Chatrchyan et al. Search for anomalous t t-bar production in the highly-boosted all-hadronic final state. *JHEP*, 1209:029, 2012.

[29] Serguei Chatrchyan et al. Search for heavy resonances in the W/Z-tagged dijet mass spectrum in pp collisions at 7 TeV. 2012.

[30] Serguei Chatrchyan et al. Studies of jet mass in dijet and W/Z+jet events. *JHEP*, 1305:090, 2013.

[31] Serguei Chatrchyan et al. Search for T5/3 top partners in same-sign dilepton final state. CMS-PAS-B2G-12-012 .

[32] Serguei Chatrchyan et al. Inclusive search for top partners. CMS-PAS-B2G-12-015 .

[33] Serguei Chatrchyan et al. Inclusive search for bottom partners. CMS-PAS-B2G-12-019 .

[34] Serguei Chatrchyan et al. Search for heavy tagged dijet resonances. CMS-PAS-EXO-12-024 .

[35] Hooman Davoudiasl, Shrihari Gopalakrishna, Eduardo Ponton, and Jose Santiago. Warped 5-Dimensional Models: Phenomenological Status and Experimental Prospects. *New J. Phys.*, 12:075011, 2010.

[36] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-kt jet clustering algorithm. *JHEP*, 04:063, 2008.

[37] Matteo Cacciari, Gavin Salam, and Gregory Soyez. Fastjet 2.3 user manual. *Phys. Lett. B*, 641, 2006.

[38] Matteo Cacciari and Gavin P. Salam. Dispelling the N**3 myth for the k(t) jet-finder. *Phys. Lett. B*, 641:57–61, 2006.

[39] S. Rappoccio, W. Adam, V. Adler, B. Hegner, L. Lista, S. Lowette, et al. PAT: The CMS physics analysis toolkit. *J.Phys.Conf.Ser.*, 219:032017, 2010.

[40] SnowMass Collaboration. Snowmass on the mississippi. `http://www.snowmass2013.org/tiki-index.php`.

[41] J. Dolen, S. Rappoccio, R. Calkins, S. Chekanov, J. Conway, R. Erbacher, et al. Reconstructing top quarks at the upgraded LHC and at future accelerators. 2013.

[42] Georges Aad et al. Performance of jet substructure techniques for large-R jets in proton-proton collisions at $\sqrt{s}$ = 7 TeV using the ATLAS detector. 2013.

[43] CMS Collaboration. Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. CMS Physics Analysis Summary CMS-PAS-PFT-09-01, 2009.

[44] Serguei Chatrchyan et al. Measurement of differential top-quark pair production cross sections in pp colisions at $\sqrt{s} = 7$ TeV. *Eur.Phys.J.*, C73:2339, 2013.

[45] Frederic Deliot, Yvonne Peters, Veronica Sorin, Yvonne Peters, and Veronica Sorin. Top quark physics at the Tevatron. *International Journal of Modern Physics A*, 08:2013, 28.

[46] Serguei Chatrchyan et al. Measurement of the charge asymmetry in top-quark pair production in proton-proton collisions at $\sqrt{s} = 7$ TeV. *Phys.Lett.*, B709:28–49, 2012.

[47] Georges Aad et al. Measurement of the charge asymmetry in top quark pair production in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector. *Eur.Phys.J.*, C72:2039, 2012.

[48] A Dominguez. Cms technical design report for the pixel detector upgrade. Technical Report CERN-LHCC-2012-016. CMS-TDR-11, CERN, Geneva, Sep 2012. Additional contacts: Jeffrey Spalding, Fermilab, Jeffrey.Spalding@cern.ch Didier Contardo, Universite Claude Bernard-Lyon I, didier.claude.contardo@cern.ch.

[49] J Mans. Cms technical design report for the phase 1 upgrade of the hadron calorimeter. Technical Report CERN-LHCC-2012-015. CMS-TDR-10, CERN, Geneva, Sep 2012. Additional contact persons: Jeffrey Spalding, Fermilab, spalding@cern.ch, Didier Contardo, Universite Claude Bernard-Lyon I, contardo@cern.ch.

[50] SUNY at Buffalo Physics Department. Higgsfest. `http://www.physics.buffalo.edu/HiggsFest`.

[51] Charlotte Hsu. Higgsfest celebrates physics discovery. `http://www.buffalo.edu/ubreporter/archive/2012_11_29/higgs_fest.html`.

[52] Righteous Babe Records. The ninth ward at babeville. `http://www.babevillebuffalo.com`.

[53] Babeville and SUNY at Buffalo. Science and art cabaret. `http://www.hallwalls.org/science-art.php`.

BUDGET JUSTIFICATION

Institution : The State University of New York at Buffalo (UB)
PI : Salvatore Rappoccio

Personnel

The requested funds of $578,755 USD (for 3 years starting in 2014) would cover two (2) months of summer salary for Prof. Rappoccio per year for three (3) years ($48,626), the full salary for one (1) postdoctoral fellow, Dr. James Dolen, for three (3) years ($154,224), and salary plus tuition for two (2) graduate students, one in-state (Joshua Kaisen) and one out-of-state (Maral Alyari), totaling $91,812 in salary and $13,740 in tuition.

Travel

This research will require significant travel to CERN. Additionally, this will cover travel costs for Dr. Dolen, Mr. Kaisen and Ms. Alyari for relocation to Fermilab or CERN. Hence, the proposal requests $31,216 in travel funds for the three years of activity.

Facilities and Administration Indirect Costs

The above figures carry an indirect cost percentage of 26% for Dr. Dolen, Mr. Kaisen and Ms. Alyari as they will not be located at the UB campus. The indirect cost for Rappoccio's summer salary is 56%, because he is located at the UB Campus. The total indirect cost is $149,563.

Below is a year-by-year summary of the funding request.

Year One (1-June-2014 – 31-May-2015)

In the first year, Dr. Dolen will perform studies of the boosted-top analyses at CMS to ensure that they will be deployable in a very early timescale in Run 2.

It is expected that the current detectors will be reinserted and incorporated in global commissioning runs collecting cosmic-ray data, followed by collision data early in 2015 for the start of Run 2. The cosmic data can be used to investigate the particle-flow algorithm reconstruction by Mr. Kaisen, as well as to begin deployment of the DQM program by Ms. Alyari. The early collision data will provide the necessary information for any further tuning.

The searches outlined above will be of high interest early in Run 2, and much of the analysis work will be in preparation for the Run 2 startup. These are expected to be designated "High Priority Analyses" (HPA's) for CMS, which means they are expected to be sensitive immediately.

In addition, Dr. Dolen is responsible for testing and commissioning of the FPIX HDI modules while stationed at Fermilab.

Mr. Kaisen will continue his work on adapting the particle-flow algorithm to work in the upgraded detector geometries in the Phase 1 and Phase 2 upgrades of CMS. He is expected to finish his classes in the Spring Semester, 2014, and will move to Fermilab for

1 year, after which time he will move to CERN. At Fermilab he will partake in further testing and commissioning of the FPIX modules.

Ms. Alyari will primarily focus on data-quality monitoring of the jet reconstruction and jet substructure algorithms. She is expected to finish her classes in the Fall Semester, 2013, and will move to Fermilab for 1-2 years. She is expected to stay at Fermilab for the remainder of her graduate studies.

Since this is the year that the LHC startup will occur, the first public outreach event will be presented.

Year Two (1-June-2015 – 31-May-2016)

Since the LHC Run 2 is expected to be in operation during the second year of this proposal, it is a period of high activity. Thus, it will be necessary to have a high presence of the group at CERN, and as such, it is expected that Dr. Dolen and Mr. Kaisen will relocate there to partake in commissioning activities and in operations for the FPIX detector, as well as to be extremely plugged into the BSM searches that will be of high interest at the time. Ms. Alyari will be located at FNAL during this period. At FNAL, it is expected that she will contribute heavily via the *Remote Operations Center* (ROC) at FNAL in the jet DQM, as well as to participate in the aforementioned BSM searches.

Year Three (1-June-2016 – 31-May-2017)

It is expected that the Run 2 LHC will be fully operational at nominal luminosity in year three of this proposal, and will have collected a large fraction of the expected Run 2 collision data for physics analysis. It is also expected that between years three and four, Run 2 will end, and part of the Phase 1 upgrades of the CMS detector will be implemented. Dolen and Kaisen will be fully engaged in the pixel operation and Phase 1 upgrade preparations, as well as data analysis. Alyari will be performing DQM support and shifts, as well as data analysis.

In addition to the Phase 1 upgrade commissioning, there will be a need to further tune the PF algorithm and the jet reconstruction software to handle the newest detector, and to be ready for startup again for Run 3. The work of Mr. Kaisen and Ms. Alyari will be imperative to the successful usage of the Phase-1 upgraded detector in collisions.

Facilities, Equipment and Other Resources

Laboratory Space

The UB group has two 900 sqft labs, VME/NIM crates, several PC's, a Tektronix Logic Analyzer, several scintillator-based muon detectors with photo-multiplier tubes, and a machine shop. In addition, the UB group has extensive access to the Silicon Detector Laboratory Facility ("SiDet") at FNAL.

Center For Computational Research

UB has a large computational research center, CCR, which is a Linux-based cluster on the Open Science Grid, and also has a large GPU cluster for possible parallel processing developments.

Office

The faculty, postdoctoral fellows and graduate students all have office space at FNAL and at CERN. In addition, when at UB, there is a large working area inside of the lab of Profs. Kharchilava and Iashvili.

Postdoctoral Fellow Mentoring

One postdoctoral fellow will be funded on this project. There are extensive postdoctoral fellowship mentoring activities at UB, at the Fermi National Accelerator Laboratory (FNAL), and via the CMS Experiment at CERN. These include guidance in career paths, work/life balance discussions, and technical skill development such as writing grant proposals, etc. Specific elements are highlighted below.

- **University at Buffalo (UB)**
  - The UB Office of Postdoctoral Scholars offers diverse services for postdoctoral fellows, including the *"Postdoc Survival Skills Workshops"*, targeted seminars and symposia for postdoctoral fellows, social functions, and logistical assistance.
  - The UB Physics Department offers several services to our postdoctoral fellows, including a biweekly Journal Club for particle physics and cosmology, weekly seminars and colloquia, and weekly social functions inside the department.
- **Fermi National Accelerator Laboratory (FNAL)**
  - The postdoctoral mentoring programs at FNAL are well-developed and tailored specifically for US-based researchers in particle physics. This includes the "LHC Physics Center" (LPC) at FNAL, which is a "brick-and-mortar" place for LHC researchers to be productive in a place other than CERN, as well as to provide much-needed visibility for postdoctoral fellows in a large collaboration.
  - FNAL provides a continuous stream of colloquia, seminars, workshops and conferences throughout the year where postdoctoral fellows can meet others in their field, expand their learning opportunities, and provide a forum for them to present their work on a regular basis.
- **CMS Experiment and CERN**
  - As at FNAL, the opportunities for a postdoctoral fellow at CERN and at CMS are extensive. There are also a plethora of workshops, seminars, conferences, etc, at CERN. There are also smaller weekly avenues for networking possibilities, as well as seminars for postdoctoral fellows to gain visibility for their work.
  - It is also worth pointing out that, because of the world-class nature of CERN, it often attracts very high-level members of the particle physics community on a regular basis. Such opportunities for visibility among the top-tier scientists in the world (including Nobel and Milner Prize winners, etc) are hard to understate.

In all, the postdoctoral fellow that will be supported by this proposal will have ample opportunities for professional advancement and development, as well as a myriad of opportunities for a community of peers in both professional and social settings.

Data Management Plan : Research Data

The CMS experiment is dedicated to timely dissemination of its data and procedures, in addition to documentation of results in publications and journal articles. The LHC experiments (including CMS) are world leaders in grid computing and cloud-like data storage, and the solutions that have been developed have robustly handled the many petabytes of data that have been collected. There are several "tiers" of data, which are designated by how widely they are deployed throughout the Open Science Grid (OSG) and the other LHC sites. The data "tiers" are :

- **RAW** : The raw data, collected in terms of simple detector readouts and event information. While rarely used for analysis, this is always available for all of the collected data at CMS indefinitely.
- **Reconstruction (RECO)** : The reconstructed data, which utilizes the RAW data tier and computes relevant information in higher-level objects to be used by analyzers.
- **Analysis Object Data (AOD)** : In Run 1, the RECO tier was too large to transmit throughout the OSG. Instead, a subset of the RECO was stored, the AOD tier. This was the primary tier for analysis usage, although it was more transient in nature. In Run 2, the computing model for this stage is still under development, although the functionality of the AOD tier will always be present.

These are stored at the various OSG sites of CMS as follows.

- **Tier 0** : The long-term data collection and storage facilities are located at CERN, where the experiments are, including CMS. Data are collected and stored in RAW format at the Tier 0 site.
- **Tier 1** : Subsequent to data-collection at the Tier-0 facility, the data are shipped in RAW format to several sites worldwide (including to FNAL in the US) at facilities where the reconstruction software is run. Data are processed and stored in the RECO format at the Tier-1 sites.
- **Tier 2** : For simulation of data, and for analysis, many smaller clusters are utilized throughout the world. These are typically stored locally in the AOD format. Since the simulated data are not stored locally at the Tier 0 site, there is one "custodial" Tier 2 which stores each generated sample centrally for the long term.

In addition to these "tiers" of the actual data collected, there are also software and documentation schema for the LHC data and analysis. The software for the reconstruction is stored and versioned locally at CERN and duplicated at FNAL, and is visible to the public [1]. This has also now been fully migrated to `github` [2]. While the data collected are initially private to CMS, there are now mechanisms in place to make the entirety of the data completely public, although this may take several years to fully realize and deploy.

[1]https://cmssdt.cern.ch/SDT/lxr/
[2]https://github.com/cms-sw/cmssw

In the meantime, there are well-defined approval procedures to ensure that the data collected by CMS are made public via documentation in webpages and journal publications, or by data-sharing projects such as `HEPDATA` [3].

[3] http://hepdata.cedar.ac.uk