P. Harris: Detecting Anomalies with AI foundation models in CMS experiment
Novel strategies in AI are quickly being adapted to pursue a wide variety of downstream applications. In this project, we pursue the construction of foundation models for the real-time data acquisition system of the CMS experiment on the LHC. Foundation Models are the building blocks of many modern deep-learning algorithms, including Large Language Models. In this project, we will collaborate with an existing team of experts at CERN to explore the use of foundation models for discovering new physics models without embedding any assumptions about the type of new physics. The focus will on real-time applications to handle the Petabit/s dataload present in future LHC operations.
Reference: https://arxiv.org/abs/2403.07066

P. Harris: Exploring Physics motivated AI compression algorithms
Data acquisition systems at the LHC far exceed that of any other device worldwide. In the CMS experiment, the raw data value can exceed 1 Petabit per second. To process this information, we utilize custom processors known as FPGAs to process the system in real-time. This project explores physics-motivated AI strategies known to compress algorithms and embed core physics algorithms into an FPGA. Our recent focus is on Kolmogorov Arnold Networks (KAN) developed at MIT, which aim to be highly expressive, and can capture many salient physically motivated features, enabling more and better algorithms that we can run in real-time on all Petabits per second.
Ref: https://arxiv.org/abs/2401.09949 https://arxiv.org/abs/2404.19756

P. Harris: Measuring interactions of the Higgs boson
The challenge of finding new particle decays of a fundamental object can be daunting. To search for new decays, one must produce the particle, and then establish a series of rules for how to search and identify these new decays. Rules for searching these are often restrictive since they require the use of hypothetical new physics models to search for their decays. In place of that, we use AI to build these rules allowing us to generically search for any new type of physics. To that extent, we explore novel AI architecture, and training strategies, to search in Higgs boson data form the CMS experiment on the Large Hadron Collider to find novel, unprecedented decays of the Higgs boson. These decays include possible dark matter, and dark sectors decays that could profoundly change our view of the universe. Additionally, through this work, we will place bounds of fundamental parameters of the Higgs boson.
Ref: http://cds.cern.ch/record/2866501?ln=en , https://inspirehep.net/literature/2809670
https://arxiv.org/abs/2412.07033

P. Harris: Searching for Dark Matter at 40 million collisions per second
Over the last year, members at CERN have started to develop a new type of data acquisition system that will allow for the processing of all collisions at data using a novel strategy known as data scouting. In this strategy, we develop complex algorithms on simplified inputs that come from the readout of the detectors on the LHC. In this project, we will explore the possibility of novel, compelling dark matter searches using this new data-scouting system being developed at CERN. Work will involve hands-on exploration of the new system, as well as exploration of how to search for dark matter at the Large Hadron Collider.

Ref: https://cds.cern.ch/record/2816252?ln=en

E. Smith, M. Williams

CODEX-beta is a brand new small detector designed to search for long-lived particles and demonstrate the feasibility of the larger proposed CODEX-b experiment. This summer project gives you the chance to take the first data with it in a small team of physicists! The detector is currently being installed near the LHCb detector on the Large Hadron Collider (LHC) at CERN. By this summer, we will start taking preliminary data to commission the detector. Part of this process will involve: (1) working with field-gate programmable arrays (FPGAs) to read out the detector; (2) developing software in the LHCb stack to interpret the detector output, reconstruct particles etc; (3) adjusting the detector to fix hardware issues that might arise. The project can focus on one or more of these areas, depending on your specific skills.

E. Smith, M. Atenzi: Matter antimatter differences in rare B decays

Rare B decays are powerful probes of new physics, and their behaviour exhibits significant deviations from Standard Model predictions. This project will develop the theoretical work needed to perform the first measurement of the differential decay rate of the rare mode B0s -> Phi(->KK)MuMu as a function of the angular distribution of the decay products, and the decay time of the B0s meson. These observables give us unique information on matter anti-matter differences. The student will develop the open source package Flavio (https://flav-io.github.io/) to include the full set of time-dependent angular observables in a rare b decay for the first time. The student will also use statistical methods to estimate the expected sensitivity to these observables when measured using the full dataset collected by the LHCb experiment.

E.Smith, Loncar : Fault-Tolerant Neural Networks for Data Compression in High-Radiation Environments

The Large Hadron Collider (LHC) at CERN generates immense volumes of data under extreme conditions, demanding robust data acquisition and processing systems. To manage the large volume of data, sophisticated compression techniques need to be applied at the source of the data, on specialized hardware as close to the detector as possible, in a high radiation environment. This project aims to design and implement fault-tolerant neural network solutions tailored for such a demanding environment. The student will explore innovative techniques such as redundancy mechanisms, error-correcting codes, and model compression strategies to ensure resilience against radiation-induced bit flips and hardware failures. The project will involve simulating radiation effects, optimizing neural architectures for real-time performance based on radiation sensitivity metrics, and testing deployment on hardware platforms used at CERN. This work will contribute to enhancing data reliability, and advancing the integration of AI into particle physics experiments, offering the student a unique opportunity to merge cutting-edge machine learning with high-energy physics challenges.

Wang, Paus: Dark Photon search with Run 3 data in CMS

Searches for resonances at the TeV scale, predicted by beyond the SM (BSM) theories, have so far yielded null results. Therefore interest has grown in extending resonance searches to lower masses. The GeV scale is particularly interesting as a new vector boson in that mass range could be a candidate for dark matter. The resonance that decays to a pair of muons is an excellent option for finding new physics. It could be produced via the dark photon mediator in the minimal dark photon model. It could also be produced via the pseudoscalar boson from the gluon fusion process in a two-Higgs-doublet model with an extra complex scalar singlet (2HDM+S). This study will focus on a low mass dark photon search based on the parked Run~3 data that the CMS experiment has taken during the years of 2022, 2023 and 2024.

D'Alfonso, Paus: Higgs to Gamma + Hcc
The measurement of the Higgs boson's coupling to fermions is a fundamental test of the Standard Model. With the extensive Run 2 and Run 3 datasets from the LHC, CMS has made significant progress in constraining the Higgs-charm coupling. Currently, searches for Higgs decays into charm quark pairs are being conducted in gluon fusion, VH, and ttH production modes. In this project, we aim to explore an alternative approach: photon-associated Higgs production. While its cross-section is much smaller, the presence of a photon provides a powerful handle to suppress the overwhelming fully hadronic background typical of gluon fusion. Additionally, the photon offers a means to control the trigger rate more effectively. Notably, Run 3 has introduced improved photon triggers, enabling a lower momentum threshold.
This study will leverage the newly collected CMS Run 3 data to investigate the feasibility and potential advantages of this production channel.

Eysermans, Paus: Beam-Induced Backgrounds at FCC-ee: impact on vertex detector occupancy and readout
Beam-induced backgrounds in lepton colliders originate from indirect interactions of beam particles, primarily through photon emissions or collective electromagnetic effects within the particle bunch. The rate of these interactions is strongly influenced by beam parameters such as bunch density and timing structure. While the energy of the resulting background particles is typically low, their contribution to detector occupancy—particularly in the vertex detector—can be significant. In this study, we characterize the beam background at the Z pole and evaluate its impact on occupancy across different vertex detector layers. These occupancy levels are then used to establish constraints on the readout system, with a particular emphasis on the digitization process.

Eysermans, Paus: Modeling and Validation of Beam Backgrounds at the FCC-ee
Beam-induced backgrounds at the FCC-ee arise from various physical processes, including beamstrahlung, coherent and incoherent pair production, and secondary interactions with residual gas or machine components. The rates and characteristics of these backgrounds are highly sensitive to beam dynamics, collision parameters, and the modeling of collective electromagnetic effects. A precise understanding of these processes is essential for accurately predicting the background environment that detectors will experience, which is critical for their optimization. In this study, we identify and analyze the dominant background mechanisms, evaluating the impact of modeling assumptions on background rates and spatial distributions. Our results provide key insights into the validity of current physics models and guide future improvements in background mitigation strategies.

Eysermans, Paus: Calibration of jet taggers using large data samples at the Z-pole
Jet taggers are trained on Monte Carlo (MC) data, however, these simulations do not perfectly describe the collision data. The large statistics at the Z-pole can be used to calibrate the jet tagger trained on ZH MC data. This can provide a calibrated ZH sample which is more realistic and give higher confidence to the outputs of the jet tagger. In this study, we will consider a Z-pole MC dataset and learn a correction with optimal transport methods to match "real" Z-pole data which will be obtained using an alternative MC generator for now. We will also learn a correction from ZH MC data to the Z MC data to cope with the different kinematic regime. Finally, these corrections will be combined and will allow us to evaluate the tagger performance on the new ZH sample.

D'Alfonso, Paus: MAPS testing
Monolithic Active Pixel Sensors (MAPS), which integrate silicon diodes and readout circuitry within the same pixels and can be manufactured using standard CMOS processes, have emerged as a promising technology for high-granularity detectors with minimal material budget. We are currently exploring this technology for the FCC-ee vertex detectors. The planned testing campaign includes: impedance scans, digital scans, noisy-pixel scans, threshold scans, characterization with radioactive sources. These tests will be conducted using a wafer probing station at CERN, which is shared with our MIT colleagues who are investigating similar technologies for the Electron-Ion Collider (EIC). More details can be found here :https://ppc.mit.edu/archives/1163

Eysermans, Paus: Optimizing a measurement of the W boson mass at the FCC-ee
The FCC-ee will run at dedicated energy points between 157 and 163 GeV, producing more than 100M pairs of W bosons. A precise measurement of the total W pair production cross-section at threshold energies allows the extraction of the mass and width of the W boson below the MeV level due to the dependency of the cross-section on the mass. Apart from the threshold scan, a direct measurement from the decay products in the hadronic and semi-leptonic final states allows direct extraction of the W boson mass. Both complementary analyses will be explored with the analysis tools available for the FCC-ee studies using state-of-the-art Monte Carlo generators and analysis techniques.

Innocenti: R&D and characterization of bent CMOS Monolithic Active Pixel Sensors.
CMOS MAPS detectors represent the most promising technology for realizing high-accuracy, low-material tracking detectors. "Stitched" Monolithic Active Pixel Sensors thinned down to 20–40μm and bent to form truly cylindrical half-barrels can be indeed used to build the next generation of vertex detectors. Due to their very limited dissipation, they do not require liquid-based cooling. Therefore, they provide a way to build tracking layers with minimal material budget and unprecedented tracking resolution. The MITHIG group is playing a leading role in the R&D and characterization of such technology, which was initially explored for the upgrade of the Inner Tracking System of the ALICE experiment at CERN, and that will be used for the Silicon Vertex Tracker of the future ePIC experiment at the EIC and in future experimental facilities. The students involved in this project will develop the readout

software and the test of the first readout data acquisition chain, working in close collaboration with the engineers and scientists from the MITHIG group currently based at CERN.

Innocenti: Artificial intelligence on integrated circuits for the next generation of silicon pixel detectors

The design of future large-area detectors based on Monolithic Active Pixel Sensors (MAPS), such as future upgrades of the ePIC detectors at the Electron-Ion Collider (EIC) or at the Future Circular Collider at CERN, faces challenges due to the need for processing the huge data flow produced by billions of active channels. While MAPS technology significantly reduces the amount of material introduced into the detector region, the data flow produced by these detectors remains overwhelmingly large. Detectors built using current MAPS technology still require an extensive number of cables and links to transport signals to dedicated processing units where track reconstruction is performed. Additionally, the readout logic embedded within MAPS can generate a substantial amount of heat, which might necessitate the use of traditional liquid-cooling techniques, significantly increasing the material budget. To address this technological challenge, our group is developing advanced data processing and reduction directly “on-chip,” significantly minimizing the data volume at the source. MAPS sensor boosted with artificial intelligence will pave the way for the development of MAPS-based detectors with significantly improved reconstruction performance compared to current systems.

Innocenti: Constraining nuclear PDFs at low-x with the first measurements of the D0 production in ultra peripheral heavy ion collisions with CMS

The ultimate physics goal of this project is to characterize the nature and properties of nuclear matter at low-x, and unveil the emergence of new regimes of QCD at low-x by performing the first measurements of ``open'' heavy-flavor photonuclear production in UPCs at the LHC. This project relies on the unique theoretical and experimental properties of these observables, and on an experimental program that enables their measurements with the CMS detector at the LHC and exploits new hardware trigger algorithms. The activity includes the development of Deep-Neural Network (DNN) tools to maximize the rapidity and the \pt-coverage of the studies, and minimize the contamination from beam-gas and hadronic collisions. The student will be embedded in one of the ongoing analyses that CMS will present during the summer/fall conferences.