A3D3 All-Hands Annual Meeting

Report of Contributions

Contribution ID: 4   Type: **not specified**

Real-Time Compression of CMS Detector Data Using Conditional Autoencoders

Saturday, 16 August 2025 16:45 (10 minutes)

Session Classification: Poster Session | Reception

Contribution ID: **5**   Type: **not specified**

GELATO: A Generic Event-Level Anomalous Trigger Option for ATLAS

Saturday, 16 August 2025 15:20 (15 minutes)

The absence of beyond-Standard-Model physics discoveries at the LHC suggests that new physics may evade conventional trigger strategies. The existing ATLAS triggers are required to control data collection rates with high energy thresholds and target signal topologies specific to only certain models. Unsupervised machine learning enables the use of anomaly detection, presenting a unique model-agnostic way to search for anomalous signatures that deviate from Standard Model expectations. We present a new trigger sequence using fast anomaly detection algorithms in both the hardware and software triggers implemented for ATLAS Run-3 data-taking. The design and performance of the triggers will be described along with their integration and commissioning strategy with an emphasis on rate stability and operational robustness. First results from analysis of data collected through this new trigger stream, focusing on validating the trigger response, will be shown. This first anomaly detection trigger for ATLAS provides a framework for future machine learning implementations in the trigger system. The approach offers potential for novel sensitivity to a broad spectrum of new physics signatures in Run-3 and beyond.

Author: ADDEPALLI, Sagar (SLAC National Accelerator Laboratory (US))

Presenter: ADDEPALLI, Sagar (SLAC National Accelerator Laboratory (US))

Session Classification: Plenary

Contribution ID: **6**   Type: **not specified**

Applying multimodal learning to Classify transient Detections Early (AppleCiDEr) I: Data set, methods, and infrastructure

Saturday, 16 August 2025 11:30 (15 minutes)

Modern time-domain surveys like the Zwicky Transient Facility (ZTF) and the Legacy Survey of Space and Time (LSST) generate hundreds of thousands to millions of alerts, demanding automatic, unified classification of transients and variable stars for efficient follow-up. We present Apple-CiDEr, a novel framework that integrates four key data modalities (photometry, image cutouts, metadata, and spectra) to overcome limitations of single-modality classification approaches. Our architecture introduces (i) two transformer encoders for photometry, (ii) a multimodal convolutional neural network (CNN) with domain-specialized metadata towers and Mixture-of-Experts fusion for combining metadata and images, and (iii) a CNN for spectra classification. Training on ˜ 30,000 real ZTF alerts, AppleCiDEr achieves high accuracy, allowing early identification.

Authors: JUNELL, Alexandra (University of Minnesota); SASLI, Argyro

Co-authors: LE CALLOCH, Antoine (University of Minnesota); WOLD, Avery; BORDER, Benny (University of Minnesota); HALE, David (Unknown); NUNES, Felipe Fontinele (University of Minnesota); SOLLERMAN, Jesper; PURDUM, Josiah; KASLIWAL, Mansi M.; XU, Maojie (University of Minnesota); RIZHKO, Mariia; Prof. GRAHAM, Matthew (California Institute of Technology); COUGHLIN, Michael (University of Minnesota); REHEMTULLA, Nabeel; SRAVAN, Niharika; WU, Shaowei (University of Minnesota); GROOM, Steven L.; CHAUDHARY, Sushant Sharma; JEGOU DU LAZ, Theophile (Astronomy Administration); QIN, Yu-Jing

Presenter: SASLI, Argyro

Session Classification: Plenary

Contribution ID: **9**   Type: **not specified**

Hardware and Algorithm Co-development (HAC) Overview

Saturday, 16 August 2025 10:00 (15 minutes)

This talk will provide an overview of HAC's achievements over the past year and briefly introduce some ongoing projects.

Presenter: LI, Pan

Session Classification: Plenary

Contribution ID: 10   Type: **not specified**

Multi-Messenger Astrophysics (MMA) Overview

Saturday, 16 August 2025 10:15 (15 minutes)

I'll give an overview with highlights of A3D3 MMA activities.

Presenter: SCHOLBERG, Kate

Session Classification: Plenary

Contribution ID: **11**   Type: **not specified**

Neuroscience (Neuro) Overview

Saturday, 16 August 2025 10:30 (15 minutes)

Update/overview presentation on NeuroAI and Neuroscience developments and achievements in the past year.

Presenter: SHLIZERMAN, Eli

Session Classification: Plenary

Contribution ID: **12**　　　　　　　　　　　　　　　　　　Type: **not specified**

High Energy Physics (HEP) Overview

Saturday, 16 August 2025 10:45 (15 minutes)

We present an overview of current and planned High Energy Physics research activities in A3D3, driven by real-time machine learning. We report the first deployment of ML-based anomaly detection at the Level-1 trigger in both CMS and ATLAS, realized through the AXOL1TL ("Anomaly eXtraction L1 Trigger Lightweight") and GELATO ("Generic Event-Level Anomalous Trigger Option") algorithms, respectively. We also highlight advances in the CMS SONIC framework ("Services for Optimized Network Inference on Coprocessors"), enabling inference-as-a-service across heterogeneous computing resources. Finally, we outline ongoing and future developments of next-generation ML triggers and supporting infrastructure for the High-Luminosity LHC, positioning CMS and ATLAS at the forefront of AI-driven discovery.

Presenter: QUINNAN, Melissa (Univ. of California San Diego (US))

Session Classification: Plenary

Contribution ID: **14**   Type: **not specified**

Machine Vision for Video-Based Material Strain Extraction

Saturday, 16 August 2025 15:35 (15 minutes)

To advance the search for 0⬚⬚⬚ decay, the LEGEND-1000 experiment will require scaling-up from its predecessor, LEGEND-200, the cryostat in particular containing a copper reentrant tube (RT) in order to create separate volumes containing underground versus atmospheric argon. As the thinnest part of the RT will only be ˜3 mm thick, small-scale pressure and strain testing is underway to confirm structural simulations. During the course of these strain tests, video recordings of the testing process were taken and we opted to use machine learning (ML) techniques to perform a frame-by-frame analysis of the footage, extracting the outline of the test cylinders and thereby tracking material deformation with time and pressure.

Meta's computer vision software Segment Anything Model 2 (SAM2) was applied to video footage of two of the pressure tests; RT diameters were extracted at multiple z-levels from the masks produced by SAM2. Optical character recognition (OCR) software was used to extract timestamps from the videos, allowing for the synchronization of RT deformation and pressure. Following a physics-based analysis, the yield strength of the copper was found to be consistent with expectations given the material and tube geometry. This demonstrates the effectiveness of non-contact ML-based strain analysis for experiment validation and opens the door for future applications in instrumentation.

Author: SIMONAITIS-BOYD, Sonata (University of California San Diego)

Co-authors: LEDER, Alexander Friedrich (Los Alamos National Laboratory); LI, Aobo (University of California San Diego)

Presenter: SIMONAITIS-BOYD, Sonata (University of California San Diego)

Session Classification: Plenary

Contribution ID: **15**   Type: **not specified**

Spatial Permutation-Invariant Neural Transformer for Consistent Intracortical Motor Decoding

Saturday, 16 August 2025 11:45 (15 minutes)

Intracortical brain-computer interfaces (iBCIs) aim to decode behavior from neural population activity, enabling individuals with motor impairments to restore motor functions and communication abilities. A central challenge in the long-term deployment of iBCIs is the nonstationarity of neural recordings, where instability of electrode recordings alters the composition and tuning of the recorded neural population across sessions. Existing approaches attempt to address this issue by explicit alignment techniques; however, they rely on fixed neural identities and require test-time labels and parameter updates, limiting their ability to generalize across sessions and imposing a computational burden during deployment. In this work, we introduce SPINT - a Spatial Permutation-Invariant Neural Transformer framework for behavioral decoding that operates directly on unordered sets of neural units. Central to our approach is a novel context-dependent positional embedding scheme that infers unit-specific identities dynamically, enabling flexible generalization across recording sessions. Our model supports inference on variable-size neural populations and allows few-shot, gradient-free adaptation using a small amount of unlabeled data from the new session. To further promote robustness to population variability, we introduce dynamic channel dropout, a regularization method for iBCI applications by simulating shifts in population composition during training. We evaluate our approach on three motor decoding tasks from the FALCON Benchmark, comprising multi-session datasets from human and non-human primates. Our approach demonstrates robust cross-session generalization, outperforming existing zero-shot and few-shot unsupervised baselines while eliminating the need for test-time alignment and fine-tuning. Our work contributes an initial step toward a flexible and practical framework for robust, scalable neural decoding in long-term iBCI applications.

Author: LE, Trung

Co-authors: Dr FANG, Hao; LI, Jingyuan; NGUYEN, Tung; MI, Lu; ORSBORN, Amy; SÜMBÜL, Uygar; SHLIZERMAN, Eli

Presenter: Dr FANG, Hao

Session Classification: Plenary

Contribution ID: **16**   Type: **not specified**

Non-Sampling Parameter Estimation for Gravitational Wave Physics

Saturday, 16 August 2025 16:45 (1h 30m)

Fast and accurate parameter estimation of gravitational wave (GW) signals is crucial in multi-messenger astrophysics. These signals are the first to arrive, requiring prompt analysis of the merger properties. However, extracting these parameters from observed binary mergers from GW detectors remains a computational bottleneck. Current approaches, such as Markov-Chain Monte Carlo (MCMC) methods and Likelihood-Free Inference (LFI), offer robust posterior estimates but rely on sampling techniques. We propose an end-to-end pipeline based on Structured State-Space Models (S4) that directly regresses both source parameters and their uncertainties from raw time series data, bypassing the need for sampling. We validate and compare our pipeline to baselines using toy models—damped harmonic oscillators and sine-Gaussian pulses. We demonstrate its applicability to realistic binary neutron star (BNS) merger signals using simulated data injected into background taken from LIGO's third observation run. BNS merger signals are particularly interesting due to their electromagnetic counterparts. Our method simplifies the inference pipeline while achieving accuracy comparable to that of existing techniques.

Authors: REISSEL, Christina (Massachusetts Inst. of Technology (US)); YOON, Kyungseop (Massachusetts Institute of Technology)

Co-author: HARRIS, Philip Coleman (Massachusetts Inst. of Technology (US))

Presenter: YOON, Kyungseop (Massachusetts Institute of Technology)

Session Classification: Poster Session | Reception

Contribution ID: **17**   Type: **not specified**

Incorporating Inelasticity Reconstruction into Neutrino Mass Ordering Studies with IceCube

Saturday, 16 August 2025 16:45 (1h 30m)

Atmospheric muon neutrinos and antineutrinos passing through the Earth experience matter effect induced oscillations, due to the interior structure of the Earth, which only affect neutrinos or antineutrinos depending on the true neutrino mass ordering (NMO). By leveraging the fact that more neutrinos are expected to be detected than antineutrinos in IceCube DeepCore, the detector can be used to probe the NMO. However, the fact that the mean inelasticity of neutrinos and antineutrinos are different has not yet been exploited to statistically separate neutrinos and antineutrinos for IceCube DeepCore, and could be used for the IceCube Upgrade. To this end, new inelasticity reconstructions were developed using two dimensional convolutional neural networks along with a model aggregating boosted decision tree for DeepCore and a graph neural network for the Upgrade. This poster will show how these reconstruction algorithms were developed and their performance. We then use these new inelasticity reconstructions as a fourth binning variable and calculate new sensitivities to determine how much of an impact this new reconstruction could have on the determination of the NMO with IceCube.

Author: PETERSON, Josh

Co-author: JACQUART, Marc

Presenter: PETERSON, Josh

Session Classification: Poster Session | Reception

Contribution ID: **18** Type: **not specified**

StreamFlex: Leveraging NoC and DFX for Scalable Dataflow Acceleration on AMD Versal V80 FPGAs

Saturday, 16 August 2025 15:50 (15 minutes)

High-level synthesis (HLS) has greatly improved the accessibility of FPGAs by enabling a faster transition from algorithmic descriptions to efficient hardware implementations. Advances in automated design space exploration (DSE) and MLIR-based compiler flows, such as ScaleHLS, have further enhanced the ability to transform high-level algorithms into optimized hardware designs. Recent research such as HIDA has extended these capabilities by automating the conversion of machine learning model graph structures into scalable dataflow architectures using HLS streams, demonstrating notable gains in both scalability and performance.

Despite this progress, current dataflow designs still depend on emulating data movement via configurable connections within the FPGA fabric. Consequently, when dataflow architectures are constructed to closely resemble the graph structure of machine learning models, they often encounter severe routing congestion as numerous nodes compete for limited connectivity resources. To mitigate this, the dataflow tool must substantially transform the original dataflow representation to fit the constraints of the more traditional von Neumann computing model. However, this transformation process introduces bottlenecks that ultimately limit both the performance and scalability of the resulting hardware designs.

To overcome these limitations, we introduce StreamFlex, a new framework that fully leverages the advanced Network-on-Chip (NoC) capabilities of the AMD Versal V80 FPGA. Rather than hard-wiring dataflow designs onto the FPGA fabric, StreamFlex uses the native NoC to efficiently route data between hardware nodes, significantly reducing the need for aggressive abstraction-level transformations. Additionally, StreamFlex utilizes dynamic function exchange (DFX) to enable runtime reconfiguration of hardware blocks based on the active regions of the dataflow graph, maximizing hardware utilization and adaptability. This approach aims to closes the gap between flexibility and performance, making FPGA acceleration practical for a wide range of applications, including machine learning, cryptography, and high-performance computing, while lowering barriers to adoption for developers.

Author: JUN, Gregory

Presenter: JUN, Gregory

Session Classification: Plenary

Contribution ID: **19**   Type: **not specified**

Supernova neutrino detection with COHERENT at the SNS

Saturday, 16 August 2025 16:45 (1h 30m)

Observation of neutrinos from the next galactic core collapse supernova will provide insights on numerous questions in physics. There are a variety of middle- to large-scale neutrino detectors currently online that will be sensitive to these neutrinos, but a better observation can be made with more detectors and varied detection channels. The COHERENT collaboration operates a variety of low energy neutrino detectors with different target nuclei in the basement of the Spallation Neutron Source at Oak Ridge National Laboratory to observe coherent elastic neutrino nucleus scattering (CEvNS) and to measure other neutrino cross sections. Though the largest of these detectors currently designed are at the tonne scale, they may be sensitive to all flavors of neutrinos from a core collapse through the neutral current CEvNS process. This poster discusses the sensitivity of COHERENT detector systems to neutrinos from core collapse and the opportunity for FPGA based online ML algorithms for triggering and early warning.

Author: QUEEN, Joshua

Presenter: QUEEN, Joshua

Session Classification: Poster Session | Reception

Contribution ID: **20**   Type: **not specified**

Multi-Messenger Astrophysics

I'll give an overview with highlights of A3D3 MMA activities.

Author: SCHOLBERG, Kate

Presenter: SCHOLBERG, Kate

Session Classification: Plenary

Contribution ID: **21**   Type: **not specified**

Super Neural Architecture Codesign Package (SNAC-Pack)

Saturday, 16 August 2025 12:00 (15 minutes)

Machine learning is a critical tool for analysis and decision making across a wide range of scientific domains, from particle physics to materials science. However, the deployment of neural networks in resource constrained environments, such as the Level-1 Trigger and edge devices, remains a significant challenge. This often requires specialized expertise in both neural architecture design and hardware optimization. To address this challenge, we introduce the Super Neural Architecture Codesign Package (SNAC-Pack), an integrated framework that automates the discovery and optimization of neural network architectures specifically tailored for hardware deployment. SNAC-Pack combines two powerful tools: Neural Architecture Codesign (NAC), which performs a two stage neural architecture search for optimal models, and the Resource Utilization and Latency Estimator (rule4ml), which predicts the resource utilization of an architecture when implemented on an FPGA. SNAC-Pack streamlines the neural architecture design process by enabling researchers to automatically explore diverse architectures optimized for both task performance and hardware efficiency. By providing quick estimates of resource utilization and latency without requiring time consuming synthesis, SNAC-Pack accelerates the development cycle. State of the art compression techniques, such as quantization aware training and pruning, further optimize the models, resulting in architectures that are prepared to be synthesized with hls4ml and deployed to FPGA hardware.

Authors: DEMLER, Dmitri; WEITZ, Jason (University of California, San Diego)

Co-authors: HAWKS, Ben (Fermi National Accelerator Lab); DUARTE, Javier Mauricio (Univ. of California San Diego (US)); TRAN, Nhan (Fermi National Accelerator Lab. (US))

Presenter: WEITZ, Jason (University of California, San Diego)

Session Classification: Plenary

Contribution ID: **22**   Type: **not specified**

Reconstruction of boosted and resolved multi-Higgs-boson events with symmetry-preserving attention networks

Saturday, 16 August 2025 16:05 (15 minutes)

The Higgs boson's self-coupling has a significant impact on the production rate of multiple Higgs bosons. Measuring the self-coupling at the CERN LHC is crucial because any deviations from our expectations could potentially lead to new discoveries of physics beyond the standard model of particle physics. Most events are fully hadronic, meaning every Higgs boson decays to a bottom quark-antiquark pair. This introduces a combinatorial challenge known as the jet assignment problem, in which jets are assigned to Higgs boson candidates. For a given event topology, symmetry-preserving attention networks (SPA-Nets) have been introduced to address this challenge. However, the complexity of this challenge increases when considering different reconstruction topologies for each Higgs boson candidate simultaneously, i.e., two "resolved" small-radius jets each containing a cascade initiated by a bottom quark or one "boosted" large-radius jet containing a merged cascade initiated by a bottom quark-antiquark pair. In this work, we generalize the SPA-Net approach to simultaneously consider both boosted and resolved reconstruction possibilities and unambiguously interpret an event as "fully resolved," "fully boosted," or in between. We report the performance of baseline methods, the original SPA-Net approach, and our generalized version on nonresonant HH and HHH production simulated by Pythia and Delphes.

Author: LI, Haoyang (Univ. of California San Diego (US))

Co-authors: SHMAKOV, Alexander (University of California Irvine (US)); MIKKELSEN, Caden Joseph (California Institute of Technology (US)); MANTILLA SUAREZ, Cristina Ana (University of Virginia (US)); WHITESON, Daniel (University of California Irvine (US)); CHAO, Darius (Univ. of California San Diego (US)); LANDSBERG, Greg (Brown University (US)); NEWMAN, Harvey (California Institute of Technology (US)); DUARTE, Javier Mauricio (Univ. of California San Diego (US)); MITIC, Jovan (University of Belgrade (RS)); WHITE, Kaitlyn (Univ. of California San Diego (US)); STA-MENKOVIC, Marko (Brown University (US)); QUINNAN, Melissa (Univ. of California San Diego (US)); FENTON, Michael James (University of California Irvine (US)); BALDI, Pierre (Univ. of California Irvine (US))

Presenter: LI, Haoyang (Univ. of California San Diego (US))

Session Classification: Plenary

Contribution ID: **23**   Type: **not specified**

SliceFI: A Heuristic-Based Fault Injection Tool for Edge Neural Networks

Saturday, 16 August 2025 11:15 (15 minutes)

When deployed in edge applications, neural networks (NNs) undergo numerous changes to ensure they adhere to strict power, performance, and size constraints while simultaneously being robust to faults. In prior work, NN robustness is evaluated using a bit-level ranking based on how sensitive an edge NN is to a fault in a given parameter bit. Unfortunately, the fault injection campaigns used to generate such sensitivity rankings are extremely time-consuming.

SliceFI is a fault injection tool that enables efficient NN robustness analysis by performing a reduced number of model computations per fault injection. Prior to fault injection, SliceFI creates layer-wise model "slices"and then caches the expected outputs for each slice. During fault injection, SliceFI determines the sensitivity of a bit using a statistical heuristic involving the deviation from its slice's expected output. By not recomputing upstream or downstream slices, SliceFI allows designers to rapidly produce bit-level sensitivity rankings during the edge NN design process.

Authors: MEZA, Andy; WENG, Olivia; KASTNER, Ryan; TRAN, Nhan (Fermi National Accelerator Lab. (US))

Presenter: MEZA, Andy

Session Classification: Plenary

Contribution ID: **24**   Type: **not specified**

Bumblebee: a Self-Supervised Transformer for Particle Kinematics, Demonstrated on the ttbar Decay Chain

Saturday, 16 August 2025 16:45 (1h 30m)

As deep learning methods and particularly Large Language Models have shown huge promise in a variety of applications, we apply a model inspired by BERT (Bidirectional Encoder Representations from Transformers), developed by Google and utilizing the multi-headed attention mechanism, to a high energy physics problem. We focus on the process of top quark-antiquark decay reconstruction and demonstrate that the model can learn the decay chain and kinematics with high accuracy via self-supervised learning. The learned decay information can be adapted to downstream tasks such as reconstruction of mass and spin correlation observables, crucial for studying top quark entanglement and search for bound states in high energy collisions. Using final-state kinematics that would be reconstructed by the CMS detector, we tokenize, mask, and take as input into the model to produce the "next" tokens, which are the generated or truth kinematics. As a result, the model learns to effectively "translate" the kinematics measured by the detector at CMS to the true kinematics of the ttbar decay with a preliminary result of 30% improvement in the target region of 340-350 GeV. In further studies, we hope to increase the scale of this tool and explore its practical applicability for the study of other processes, as the model can easily be applied to any decay process which gives it significant potential for future studies in the high energy domain.

Authors: WILDRIDGE, Andrew (Purdue University (US)); RODGERS, Jack Patrick (Purdue University (US))

Co-authors: COLBERT, Ethan (Purdue University (US)); LIU, Mia (Purdue University); YAO, Yao (Purdue University (US))

Presenter: COLBERT, Ethan (Purdue University (US))

Session Classification: Poster Session | Reception

Contribution ID: **25**   Type: **not specified**

Machine Learning Topics in A3D3

Saturday, 16 August 2025 08:45 (1 hour)

Throughout different scientific disciplines, there is a need for machine learning models that leverage domain knowledge through inductive bias and data representations to maximize their potential. In addition, efficient machine learning implementations optimized for inference in hardware are critical for low-latency, high-throughput, or limited-resource scientific applications. In this presentation, I will review machine learning methods, like graph neural networks, transformers, and symmetry-equivariant networks, that provide ways of incorporating specialized domain knowledge, as well as effective techniques for reducing computation in neural networks, including pruning, quantization, and knowledge distillation.

Presenter: DUARTE, Javier Mauricio (Univ. of California San Diego (US))

Session Classification: Overview

Contribution ID: **26**   Type: **not specified**

Neuroscience (Neuro) Overview

Update/overview presentation on NeuroAI and Neuroscience developments and achievements in the past year.

Author: SHLIZERMAN, Eli

Presenter: SHLIZERMAN, Eli

Session Classification: Plenary

Contribution ID: **27**   Type: **not specified**

Machine Learning Topics in A3D3

Throughout different scientific disciplines, there is a need for machine learning models that leverage domain knowledge through inductive bias and data representations to maximize their potential. In addition, efficient machine learning implementations optimized for inference in hardware are critical for low-latency, high-throughput, or limited-resource scientific applications. In this presentation, I will review machine learning methods, like graph neural networks, transformers, and symmetry-equivariant networks, that provide ways of incorporating specialized domain knowledge, as well as effective techniques for reducing computation in neural networks, including pruning, quantization, and knowledge distillation.

Author: DUARTE, Javier Mauricio (Univ. of California San Diego (US))

Presenter: DUARTE, Javier Mauricio (Univ. of California San Diego (US))

Session Classification: Overview

Contribution ID: **28**   Type: **not specified**

Spatial Permutation-Invariant Neural Transformer for Consistent Intracortical Motor Decoding

Saturday, 16 August 2025 18:15 (1h 30m)

Intracortical brain-computer interfaces (iBCIs) aim to decode behavior from neural population activity, enabling individuals with motor impairments to restore motor functions and communication abilities. A central challenge in the long-term deployment of iBCIs is the nonstationarity of neural recordings, where instability of electrode recordings alters the composition and tuning of the recorded neural population across sessions. Existing approaches attempt to address this issue by explicit alignment techniques; however, they rely on fixed neural identities and require test-time labels and parameter updates, limiting their ability to generalize across sessions and imposing a computational burden during deployment. In this work, we introduce SPINT - a Spatial Permutation-Invariant Neural Transformer framework for behavioral decoding that operates directly on unordered sets of neural units. Central to our approach is a novel context-dependent positional embedding scheme that infers unit-specific identities dynamically, enabling flexible generalization across recording sessions. Our model supports inference on variable-size neural populations and allows few-shot, gradient-free adaptation using a small amount of unlabeled data from the new session. To further promote robustness to population variability, we introduce dynamic channel dropout, a regularization method for iBCI applications by simulating shifts in population composition during training. We evaluate our approach on three motor decoding tasks from the FALCON Benchmark, comprising multi-session datasets from human and non-human primates. Our approach demonstrates robust cross-session generalization, outperforming existing zero-shot and few-shot unsupervised baselines while eliminating the need for test-time alignment and fine-tuning. Our work contributes an initial step toward a flexible and practical framework for robust, scalable neural decoding in long-term iBCI applications.

Author: LE, Trung

Co-authors: ORSBORN, Amy; SHLIZERMAN, Eli; FANG, Hao; LI, Jingyuan; MI, Lu; NGUYEN, Tung; SÜMBÜL, Uygar

Presenter: FANG, Hao

Session Classification: Poster Session | Reception

Contribution ID: **29**   Type: **not specified**

Hardware and Algorithm Co-development (HAC) Overview

This talk will provide an overview of HAC's achievements over the past year and briefly introduce some ongoing projects.

Author: LI, Pan

Presenter: LI, Pan

Session Classification: Parallel

Contribution ID: **30**   Type: **not specified**

High Energy Physics (HEP) Overview

We present an overview of current and planned High Energy Physics research activities in A3D3, driven by real-time machine learning. We report the first deployment of ML-based anomaly detection at the Level-1 trigger in both CMS and ATLAS, realized through the AXOL1TL ("Anomaly eXtraction L1 Trigger Lightweight") and GELATO ("Generic Event-Level Anomalous Trigger Option") algorithms, respectively. We also highlight advances in the CMS SONIC framework ("Services for Optimized Network Inference on Coprocessors"), enabling inference-as-a-service across heterogeneous computing resources. Finally, we outline ongoing and future developments of next-generation ML triggers and supporting infrastructure for the High-Luminosity LHC, positioning CMS and ATLAS at the forefront of AI-driven discovery.

Author: QUINNAN, Melissa (Univ. of California San Diego (US))

Presenter: QUINNAN, Melissa (Univ. of California San Diego (US))

Session Classification: Parallel