Onboard AI and Federated Learning for Cosmic Ray Event Detection

Carlyn Lee, California Institute of Technology & Oligo Space, cblee@caltech.edu

Abstract

This Network Research Exhibition (NRE) proposes to demonstrate the application of Artificial Intelligence (AI) and federated learning for cosmic ray event detection. This demonstration will focus on ground-based infrastructure in preparation for flight on an Oligo Spacecraft, leveraging the cosmic ray data collected by the Cosmic-Ray Extremely Distributed Observatory (CREDO) project. A key component of this demonstration will be the integration of CosmicWatch Desktop Muon Detector units, which provide real-time cosmic ray muon detection with high temporal precision and rich sensor metadata. These compact, low-power (0.5 W) detectors utilize plastic scintillators and silicon photomultipliers (SiPM) to record cosmic-ray muons, enabling continuous data collection for AI model training and federated learning coordination.

The demonstration will showcase a complete data pipeline from detection to analysis: CosmicWatch detectors stream detection events in real-time to an Elasticsearch data infrastructure, where event data—including ADC (Analog-to-Digital Converter) values, SiPM voltages, coincidence flags, environmental sensors (temperature, pressure), and motion sensors (accelerometer, gyroscope)—are immediately indexed for processing. Federated learning algorithms will then train AI models across distributed nodes using this real-time data stream, enabling efficient classification of cosmic ray events while preserving data sovereignty across participating institutions. This end-to-end system demonstrates high-throughput data transfers and distributed processing for scientific discovery, utilizing SCinet's advanced network capabilities to coordinate real-time model updates and data synchronization across geographically distributed nodes.

This work represents a critical step toward space-based cosmic ray detection, where similar detectors will operate with limited communication bandwidth and power constraints. By demonstrating efficient edge AI processing, distributed learning, and real-time data management with ground-based CosmicWatch detectors, we validate the operational concepts and network architectures needed for future space missions. The integration of real-time detector data with federated learning workflows showcases how intelligent edge processing can reduce bandwidth requirements while enabling collaborative scientific discovery.

Goals

Our primary goal is to illustrate how AI at the edge, trained and tested with real-world cosmic ray data from projects like CREDO and real-time CosmicWatch detector streams, can enable efficient classification of cosmic ray events. Specifically, we aim to:

1. Demonstrate real-time data collection and processing: Stream cosmic ray detection events from CosmicWatch detectors directly to Elasticsearch infrastructure, processing detection data (ADC values, SiPM voltages, coincidence events, environmental parameters) with sub-second latency for immediate AI model input.

2. Enable efficient federated learning with edge detectors: Train federated learning models using real-time CosmicWatch data distributed across multiple institutions, showcasing how edge devices can participate in collaborative AI training while minimizing bandwidth requirements.

3. Validate space-ready architectures: Validate the network architectures, data protocols, and AI processing pipelines needed for future space-based cosmic ray detection, using ground-based CosmicWatch detectors as test platforms.

This work is in anticipation of future space-based detection capabilities and aligns with efforts by the Global Network Advancement Group (GNA-G) Data Intensive Sciences (DIS) Working Group, in which Oligo Space collaborates with Caltech and other institutions, to accelerate data-intensive science workflows.

Impacts

This demonstration will highlight tools in automation and AI for real-time data applications, with particular emphasis on edge-based detector integration and distributed learning systems. The integration of CosmicWatch Desktop Muon Detector units demonstrates how low-power, compact detectors can contribute to large-scale distributed scientific networks, providing high-quality cosmic ray event data with rich metadata for AI training.

This research advances the field of distributed scientific computing by demonstrating how edge detectors can seamlessly integrate into federated learning networks. Unlike traditional centralized data collection, our approach enables real-time data streaming from multiple geographically distributed detectors, allowing AI models to train on fresh, continuously updated data streams. This is particularly relevant for space missions where detectors must operate autonomously and contribute to distributed analysis networks.

SCinet's infrastructure enables streaming from multiple CosmicWatch detectors simultaneously, supporting concurrent data collection from dozens of detectors across the exhibition floor. Using SCinet's capabilities, we can demonstrate the potential of distributed edge detector networks or showcase real-time federated learning with live detector data. The network's low-latency pathways are appropriate to explore space-ready communication protocols where detector data must reach analysis systems with minimal delay. The next steps include:

(1) Deploying CosmicWatch detectors during the SC25 exhibition to demonstrate live data collection and real-time federated learning;

(2) Validating the scalability of our system to support hundreds of simultaneous detector streams;

(3) Demonstrating power-efficient edge AI processing suitable for space deployment; and

(4) Transitioning the validated ground-based system architecture to flight-ready hardware for Oligo Spacecraft deployment.

Resources

Data Sources:

CREDO.science server (https://api.credo.science/api/v2) Accessing existing cosmic ray detection data using data exporter scripts.

CosmicWatch Desktop Muon Detector v3X units streaming real-time detection events to NRP Nautilus

Central processing hub on NRP Nautilus: Caltech node (patternlab.calit2.optiputer.net, San Diego, CA) serving as federated learning server and data coordination center

CosmicWatch Detector Specifications:

USB streaming of detection events (115200 baud serial)

Detection parameters: ADC values (0-4095), SiPM voltages, coincidence flags

Environmental sensors: Temperature, pressure; Motion sensors: 3-axis accelerometer, 3-axis gyroscope

Data streaming rate: Variable based on cosmic ray flux (typically 1-10 events/minute)

Power consumption: 0.5 W (demonstrates space-ready power efficiency)

Note: CosmicWatch detectors connect via USB serial interface to local computers and do not require network VLANs

Bandwidth Needs:

Minimum 10 Gbps for high-speed data transfer of CREDO data, real-time CosmicWatch detector streams, classified events, and federated learning model updates, with a possibility for higher capacity (e.g., 100 Gbps) to demonstrate SCinet capabilities.

Real-time detector streams require low-latency pathways (<100ms) for immediate processing and federated learning coordination.

Network Connectivity:

SCinet booth connectivity to NRP Nautilus infrastructure (patternlab.calit2.optiputer.net) for federated learning coordination and data visualization

CREDO.science server (https://api.credo.science/api/v2) accessed through NRP Nautilus

2 smartphones running CREDO application. Detections upload to CREDO.science API via Wi-Fi connectivity

Data Infrastructure:

Elasticsearch cluster for real-time event indexing and storage

Kibana visualization platform for real-time detector data monitoring

Federated learning coordination server for model training across distributed nodes

This experiment will represent a single, focused demonstration of our preparatory work and capabilities leveraging CREDO data and real-time CosmicWatch detector integration.

Involved Parties

- John Graham, University of California San Diego, jjgraham@eng.ucsd.edu
- Prof. Harvey Newman, High Energy Physics, California Institute of Technology, hnewman@caltech.edu
- Prof. Spencer N.G. Axani, University of Delaware, saxani@udel.edu
- Mohammad Sada, San Diego Supercomputer Center, mfsada@sdsc.edu