Contribution ID: **205**                                                                                     Type: **Oral**

# EJFAT: Towards Intelligent Compute Destination Load Balancing

*Monday, 24 October 2022 17:40 (20 minutes)*

To increase the science rate for high data rates/volumes, JLab is partnering with ESnet for development of an AI/ML directed dynamic Compute Work Load Balancer (CWLB) of UDP streamed data. The CWLB is an FPGA featuring dynamically configurable, low fixed latency, destination switching and high throughput. The CLWB effectively provides seamless integration of edge / core computing to support direct experimental data processing for immediate use by JLab science programs and others such as the EIC as well as data centers of the future. The ESnet/JLaB FPGA Accelerated Transport (EJFAT) project is targeting near future projects requiring high throughput and low latency for both hot and cooled data for both running experiment data acquisition systems and data center use cases.

The essential function of the CWLB data plane is to redirect so designated data channel streams sharing a common data event designation to selectable destination hosts as a function of data event id, and target host ports as a function of data channel id. Thus is effected a form of hierarchical horizontal scaling at two levels; the first across compute host machines data event by data event for a type of pipe-lined processing for a series of events and secondly across ports on a compute host so that different data id channels may be assigned to different processors for parallelized further processing, e.g., reassembly, event reconstruction, physics harvesting, etc.

An EJFAT control plane running external to the CLWB and using both network and compute farm telemetry, effects AI directed and predictive resource allocation, capacity assessment, and scheduling of compute farm resources in order to dynamically reconfigure the CLWB in-situ as the operating context and conditions require.

## Significance

innovation for hierarchical horizontal scaling / load balancing

## References

## Experiment context, if any

**Primary author:**   GOODRICH, michael

**Presenter:**   GOODRICH, michael

**Session Classification:**   Track 1: Computing Technology for Physics Research