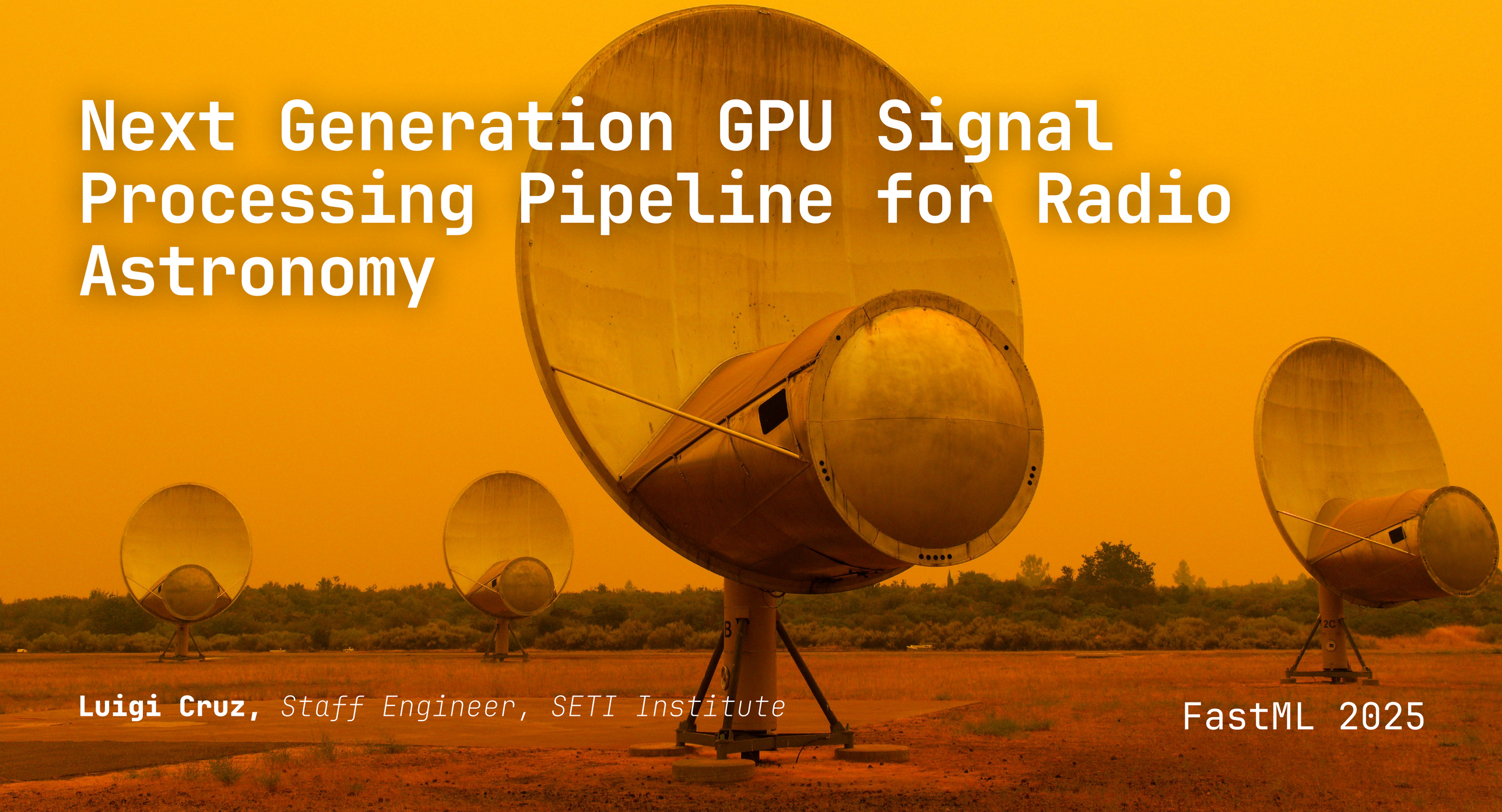
Next Generation GPU Signal Processing Pipeline for Radio Astronomy



Luigi Cruz, *Staff Engineer, SETI Institute*

FastML 2025

ALLEN TELESCOPE ARRAY

Allen Telescope Array

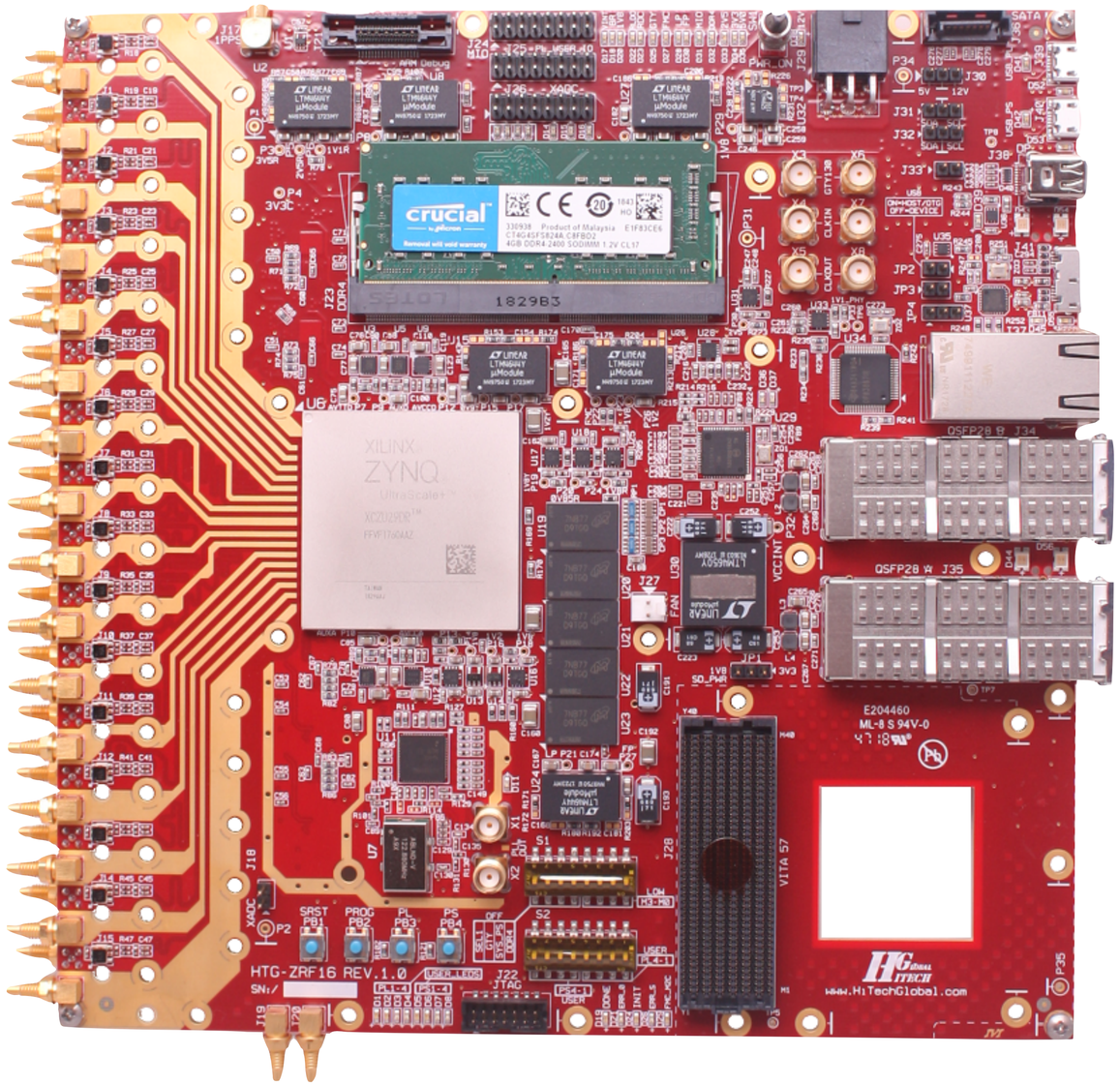
Offset Gregorian Dish

- Each of the 42 antennas has 20 feet (6.1 m) in diameter.
- Produces ~1.5 GHz of bandwidth for each polarization (~3.0 GHz in total).
- The entire telescope equates to ~84 GHz or ~1.4 Tbps at 8 bits per sample.
- Connected to the DSP Room via RF over fiber.
- Ultra-wideband reception.

Allen Telescope Array

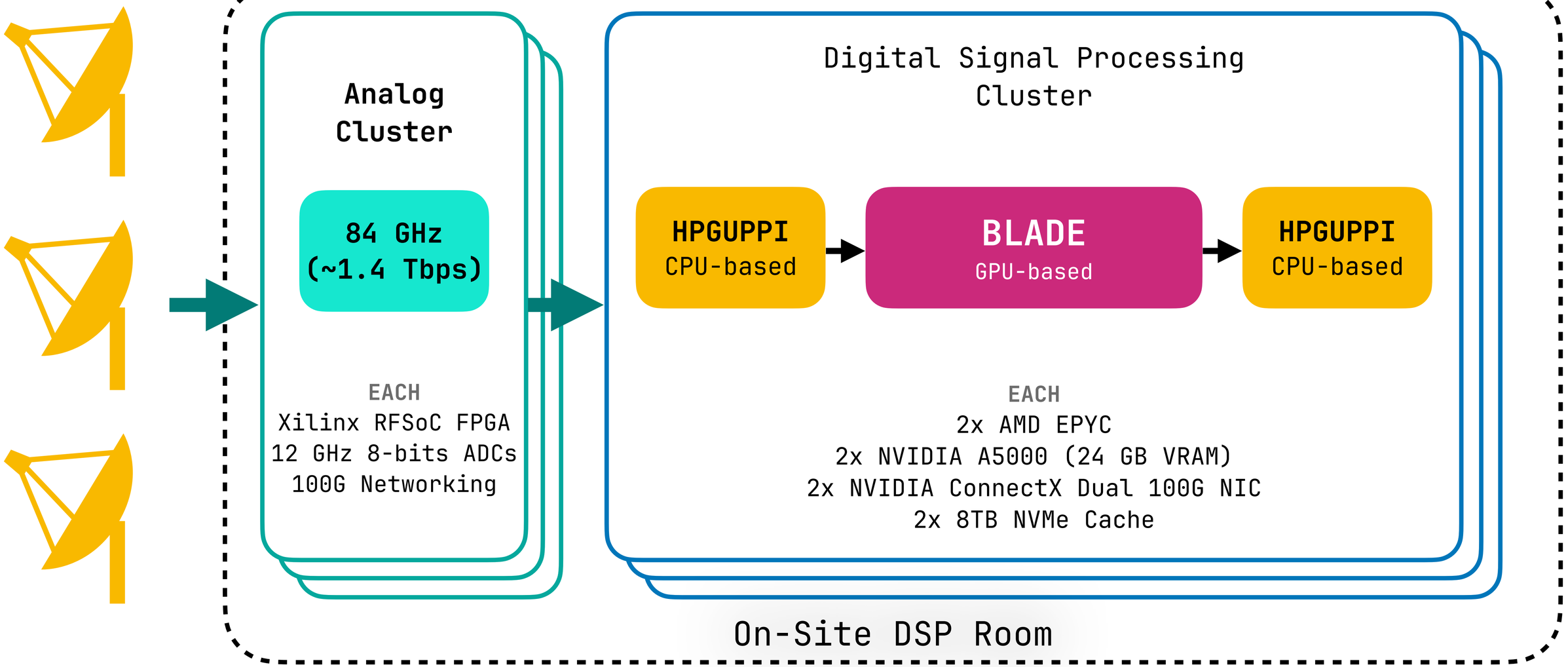
Data Acquisition

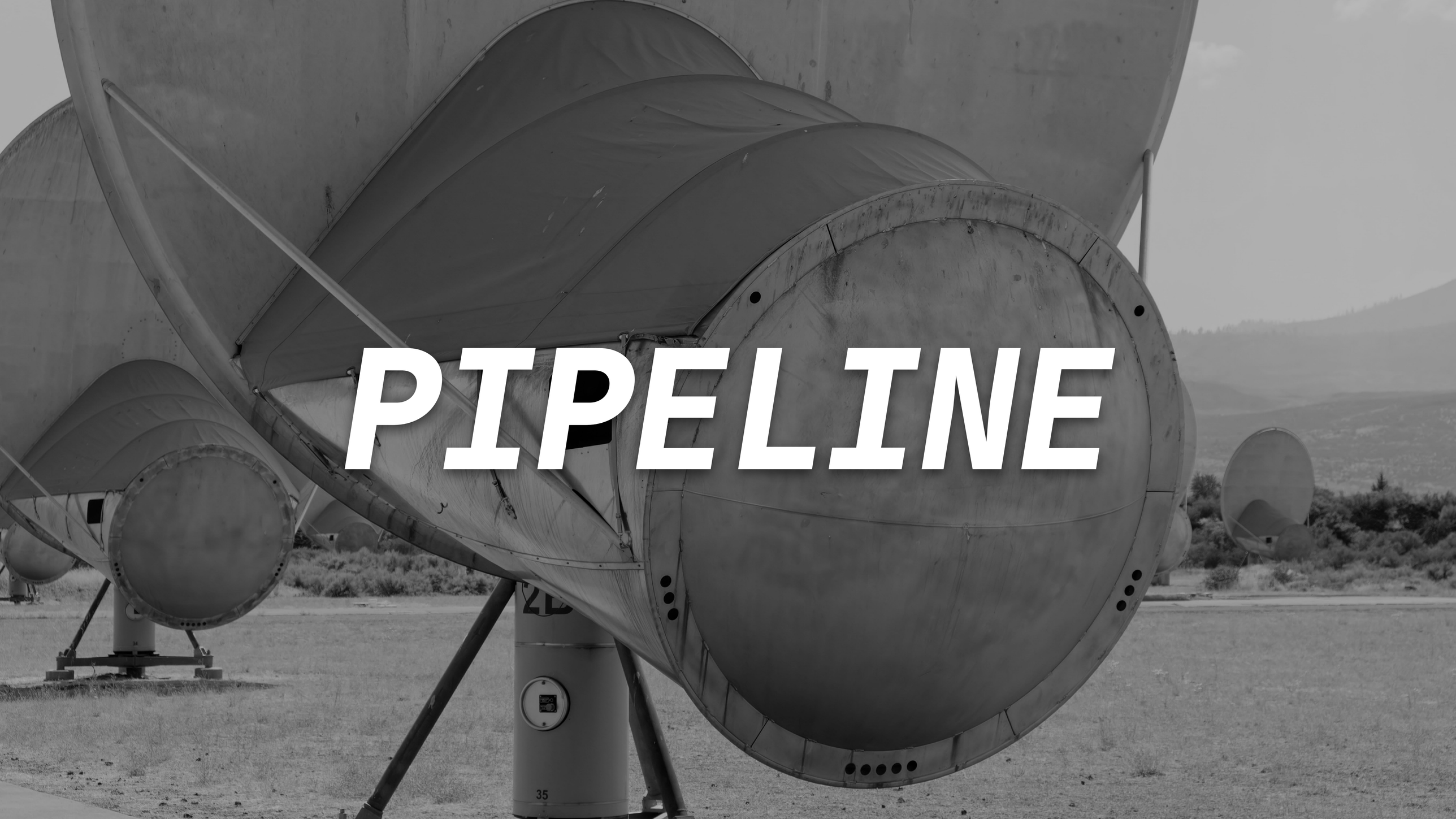
- Received radio signal is transmitted to the DSP room via RF over fiber.
- Signal is converted back to copper, pre-amplified, mixed, and distributed to the data-acquisition boards.
- Signal is digitized using RFSoC FPGA boards where it is pre-channelized, packetized, and sent over the network via 100G fiber.
- Data is received in the processing nodes.



Data Processing

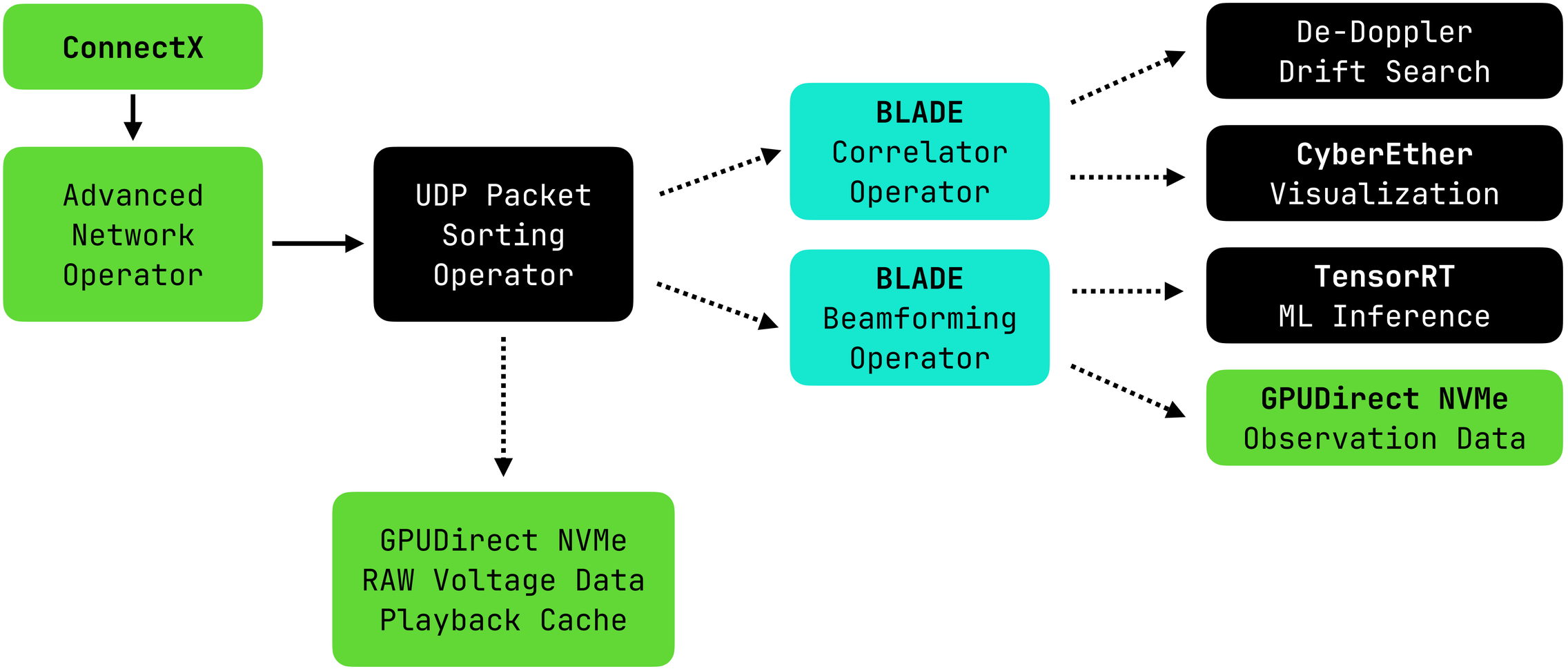
Current Pipeline

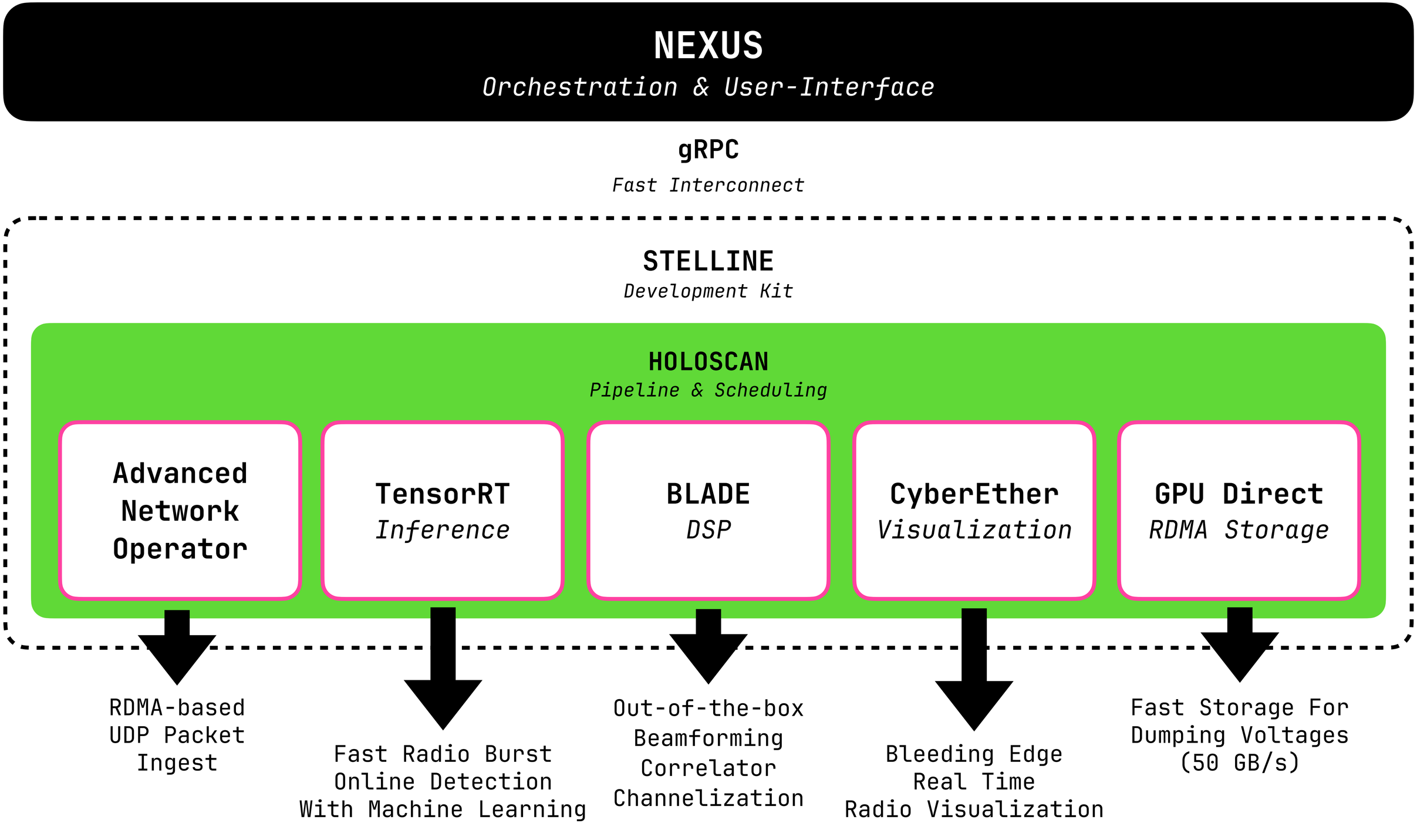



PIPELINE

Initial Goals

Extending the ATA's Capabilities with Holoscan



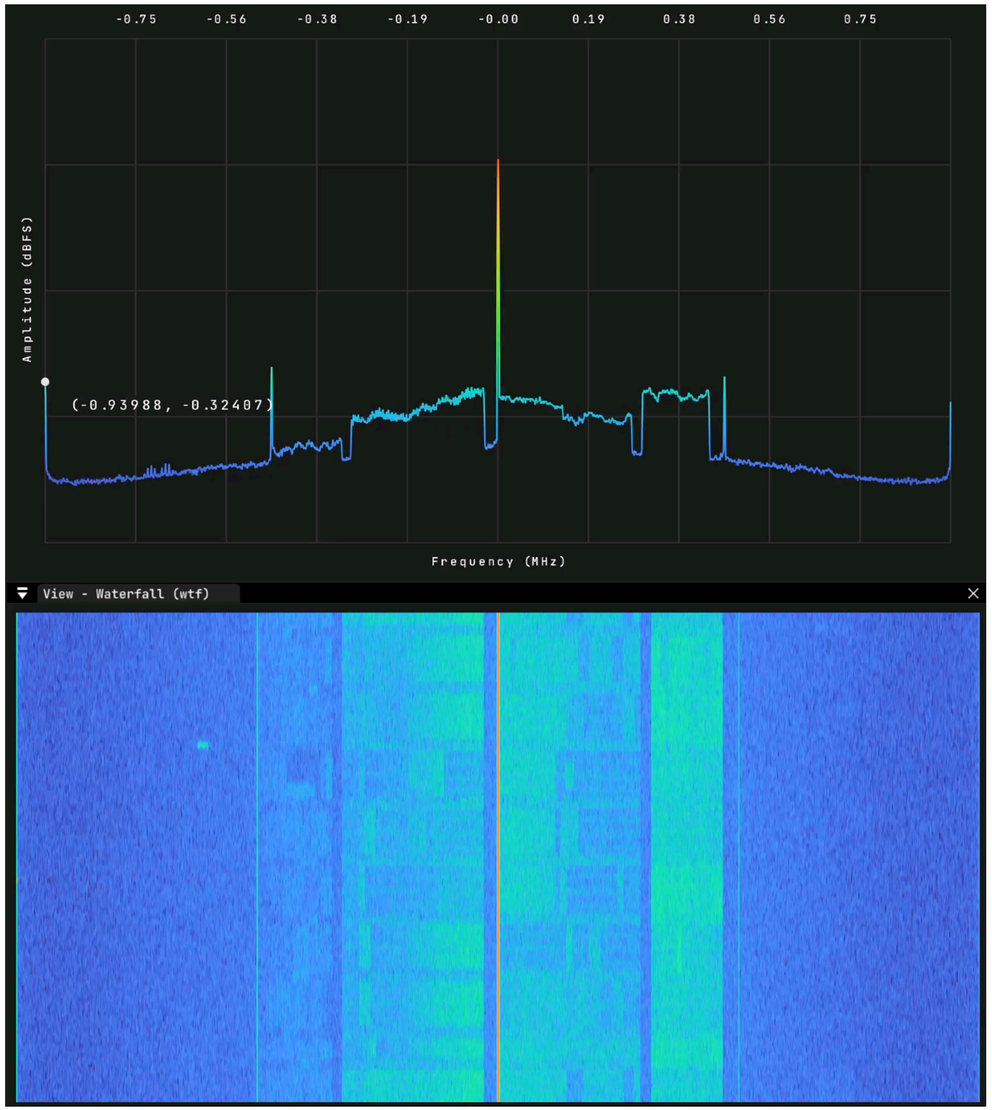

NEXUS
Orchestration & User-Interface
gRPC
Fast Interconnect
STELLINE
Development Kit
HOLOSCAN
Pipeline & Scheduling
Advanced Network Operator
TensorRT
Inference
BLADE
DSP
CyberEther
Visualization
GPU Direct
RDMA Storage
RDMA-based UDP Packet Ingest
Fast Radio Burst Online Detection With Machine Learning
Out-of-the-box Beamforming Correlator Channelization
Bleeding Edge Real Time Radio Visualization
Fast Storage For Dumping Voltages (50 GB/s)

CyberEther

Multi-Platform Hardware Accelerated Visualization

https://github.com/luigifcruz/CyberEther

- Suite of multi-platform hardware accelerated visualization tools:
  - **CyberEther**: Modular flowgraph interface for prototyping.
  - **Superluminal**: High-level Python and C++ visualization API.
- Built upon a unified set of backends:
  - **Render**: Efficiently draws the visualizations. Framework agnostic graphical API backed by Vulkan, Metal, or WebGPU.
  - **Compute**: Executes computations and uses hardware-acceleration when available including CUDA.
  - **Viewport**: Handles interactions with the host system and offers low latency remote interface.
  - **Memory**: Manages memory tensors and handles zero-copy interoperability between frameworks.
- Powerful integration with Holoscan.
- Ideal to visualize high-bandwidth spectrum data.



Stelline

Signal Processing Development Kit for Holoscan

https://github.com/luigifcruz/stelline

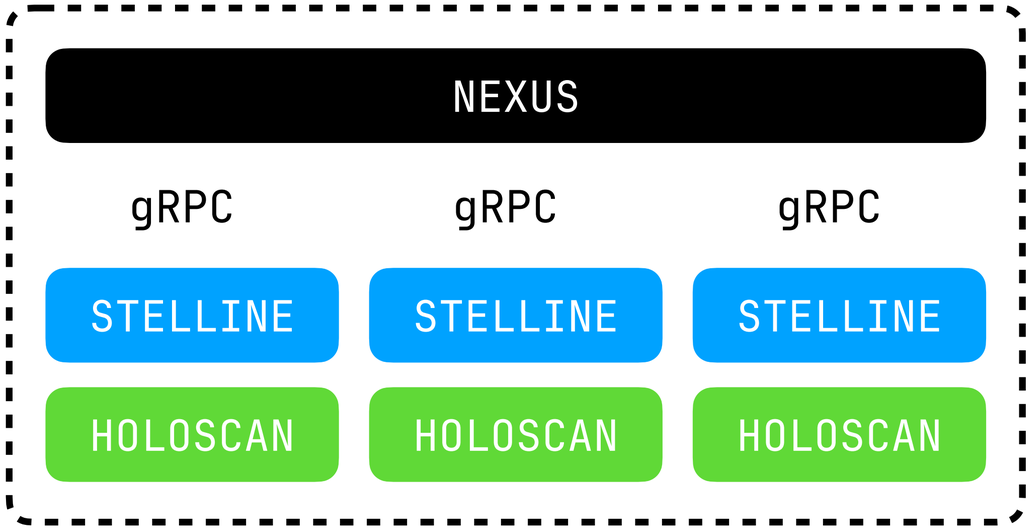
- Modular and composable real-time processing framework based on NVIDIA Holoscan.
- Aggregates custom Holoscan operators and glue code:
  - **BLADE**: Digital Signal Processing operators used for beamforming, correlation, etc.
  - **TensorRT**: Efficient inference in real-time on spectrogram data. Used by the FRBNN project.
  - **Transport**: Data reception via RDMA using the Advanced Network Operator.
  - **Filesystem**: Leverages NVIDIA GPUDirect Storage to store data to disk directly from the GPU memory. Support for HDF5 planned soon.
- Aims to replace the CPU-based pipeline at the Allen Telescope Array and other telescopes around the world.
- Customizable pipelines defined via YAML file.

NEXUS

Orchestration & User-Interface

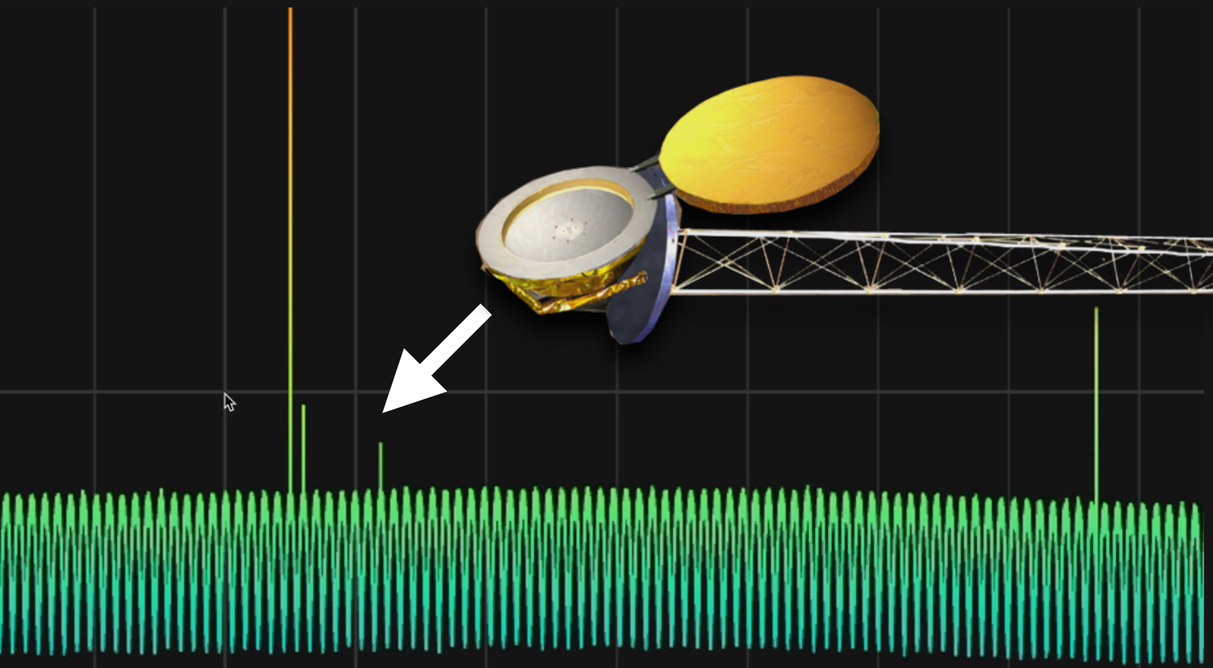
https://github.com/luigifcruz/nexus

- Orchestrates Stelline instance running in multiple machines.
  - **Metrics**: Collect and stores hardware and software metrics.
  - **Metadata**: Provides instances with observation metadata and instrument status.
  - **Instancing**: Provision, deploy, and monitor Stelline containers.
- Aims to replace the command line control interface at the Allen Telescope Array and other instruments.
- No terminal required. Modern web-based user interface.
- Written in Rust with gRPC as the communication protocol.



Deep Space Demonstration

Allen Telescope Array Holoscan Pipeline

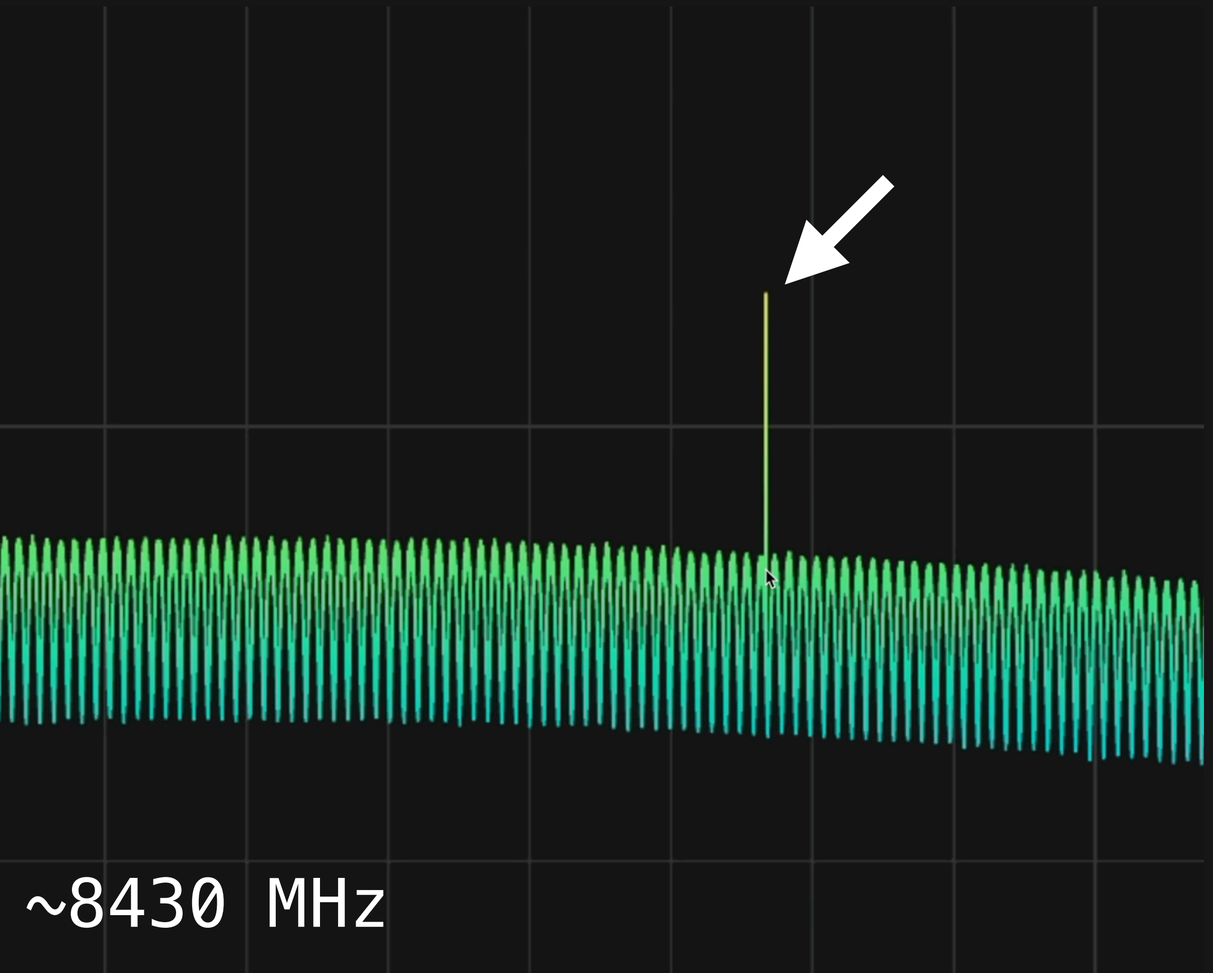


~8406 MHz

Mars Odyssey

Deep Space Demonstration

Allen Telescope Array Holoscan Pipeline

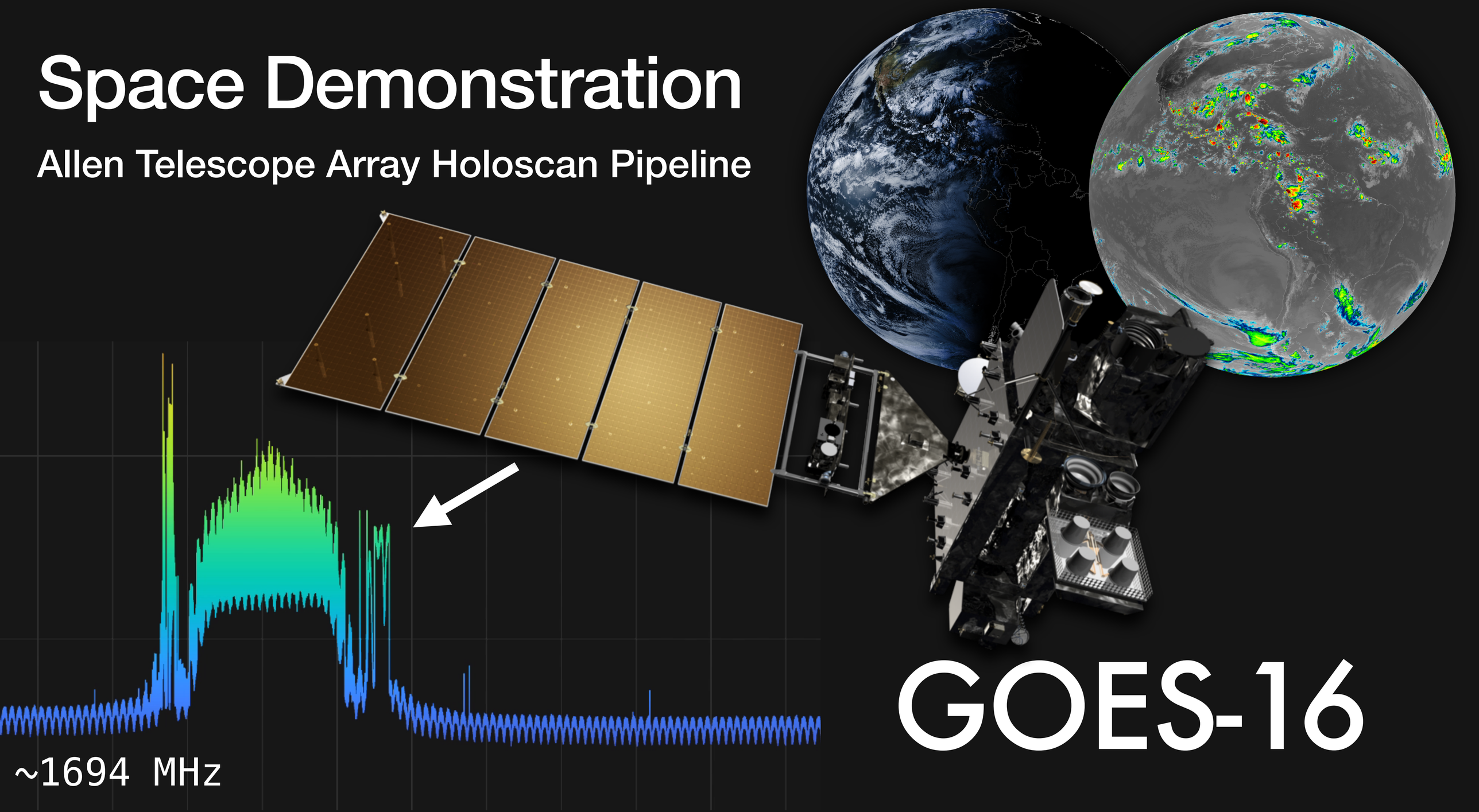


Tianwen

天问一号

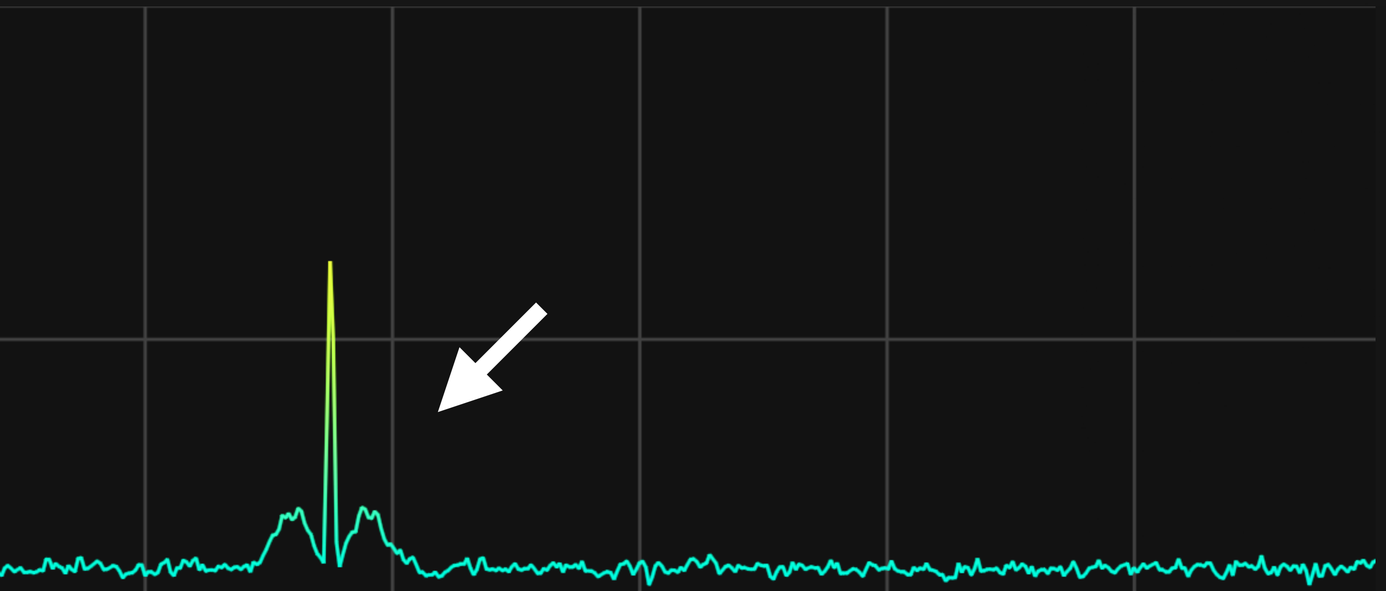
Space Demonstration

Allen Telescope Array Holoscan Pipeline



Deep Space Demonstration

Allen Telescope Array Holoscan Pipeline



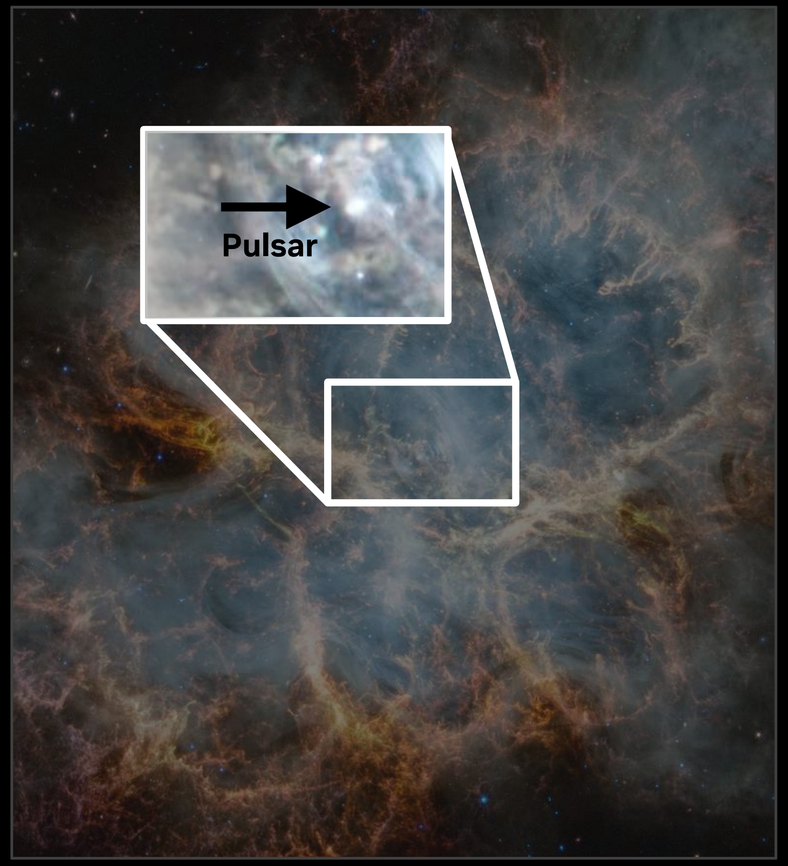
~8443.5 MHz

STEREO-A

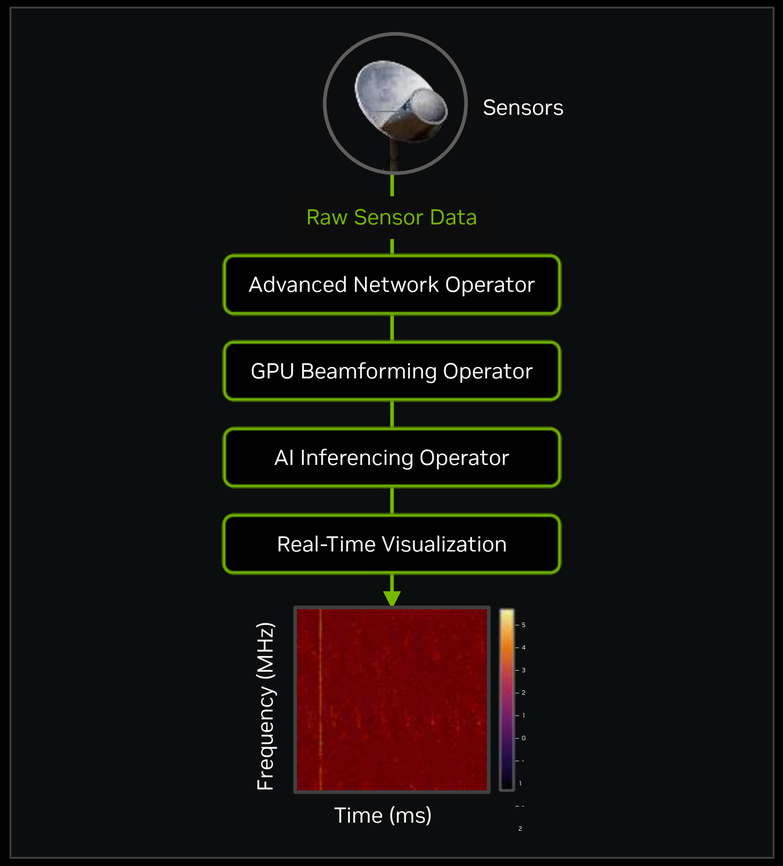

MODEL

First Real-Time Pure AI Detection of a Pulsar Using Raw Streaming Sensor Data

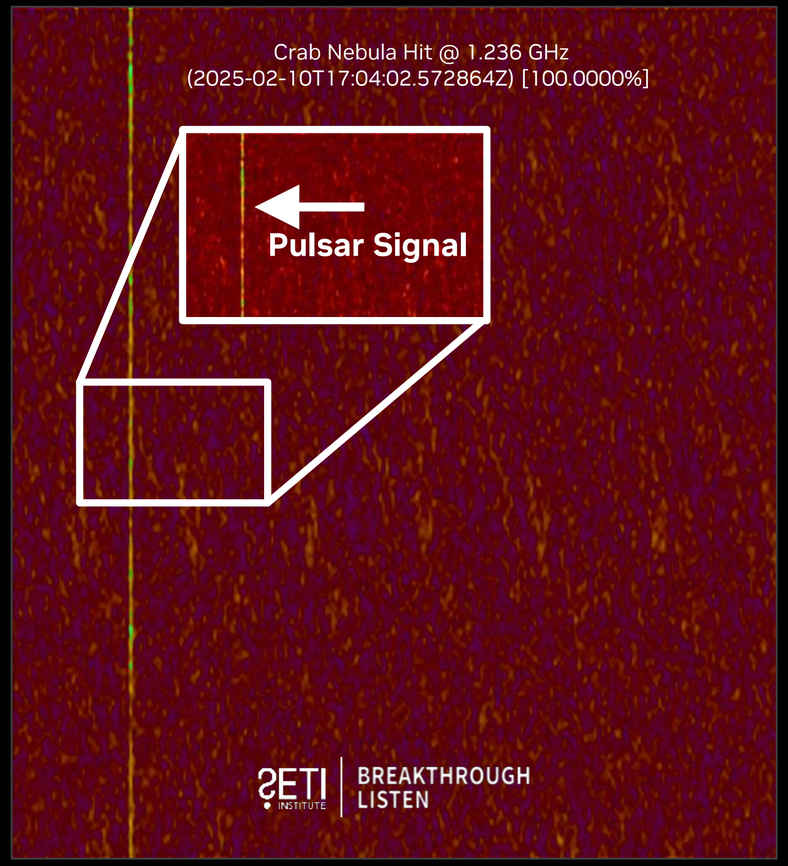
Holoscan Enables Real-Time AI-Powered Sensor Workloads at the Edge



Pulsar in the Crab Nebula



Holoscan From Beamformer to AI Model



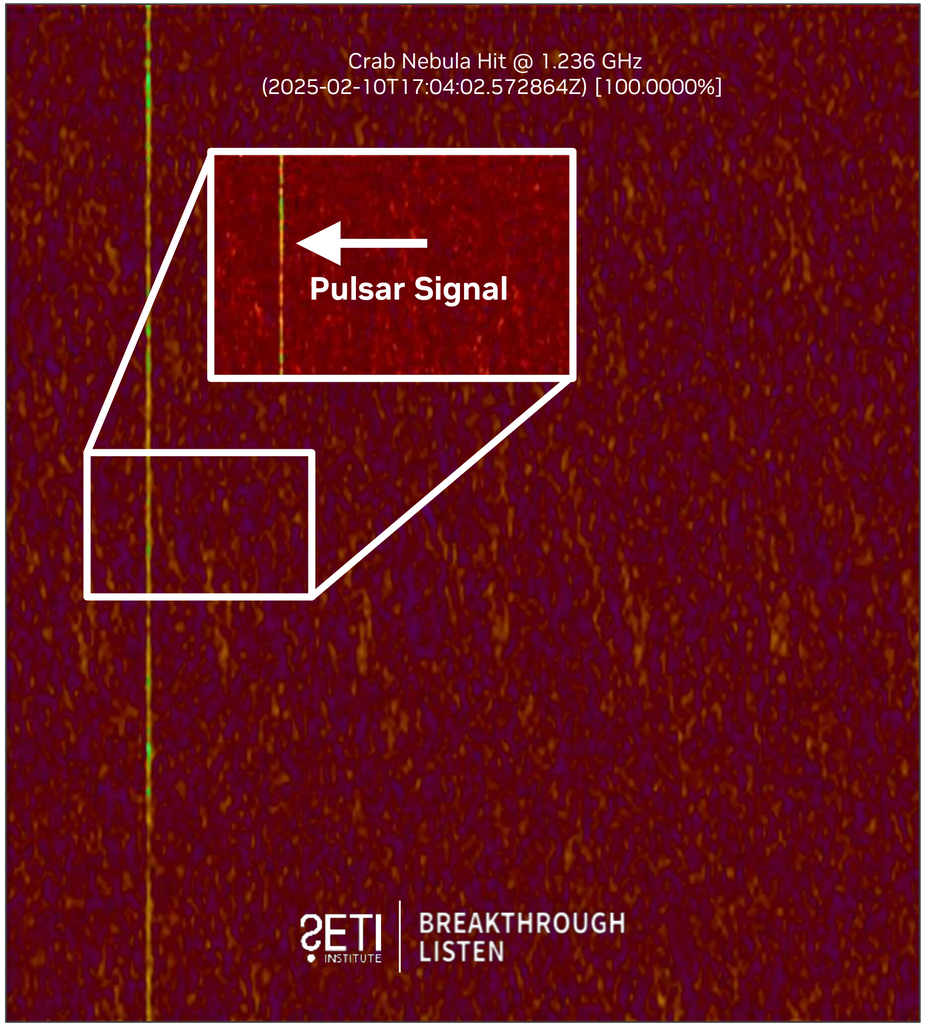
Signal Detection Beyond Noise

Fast Radio Burst Neural Network (FRBNN)

Real-Time End-to-End Deep Learning Algorithm for Fast Radio Burst Detection

https://github.com/PetchMa/frbnn

- Designed for Fast Radio Bursts (FRBs) detection.
- Lightweight model (82 MB) based on ResNet.
- Efficient TensorRT inference in real time.
- Simulation augmented training dataset:
  - Real observations from the Allen Telescope Array as base.
  - Simulated bursts synthetized via SciPy Signal and InjectFRB.
  - Resulted in 300 GB set containing 200K bursts.
- High recall rate throughout wide range of SNRs.
- Tested in a real-time setting at the Allen Telescope Array (ATA).
- Successfully identified the Crab Pulsar (PSR B0531+21) bursts.
- Paper (Ma et al. 2025) under review Astronomy & Astrophysics.



Fast Radio Burst Neural Network (FRBNN)

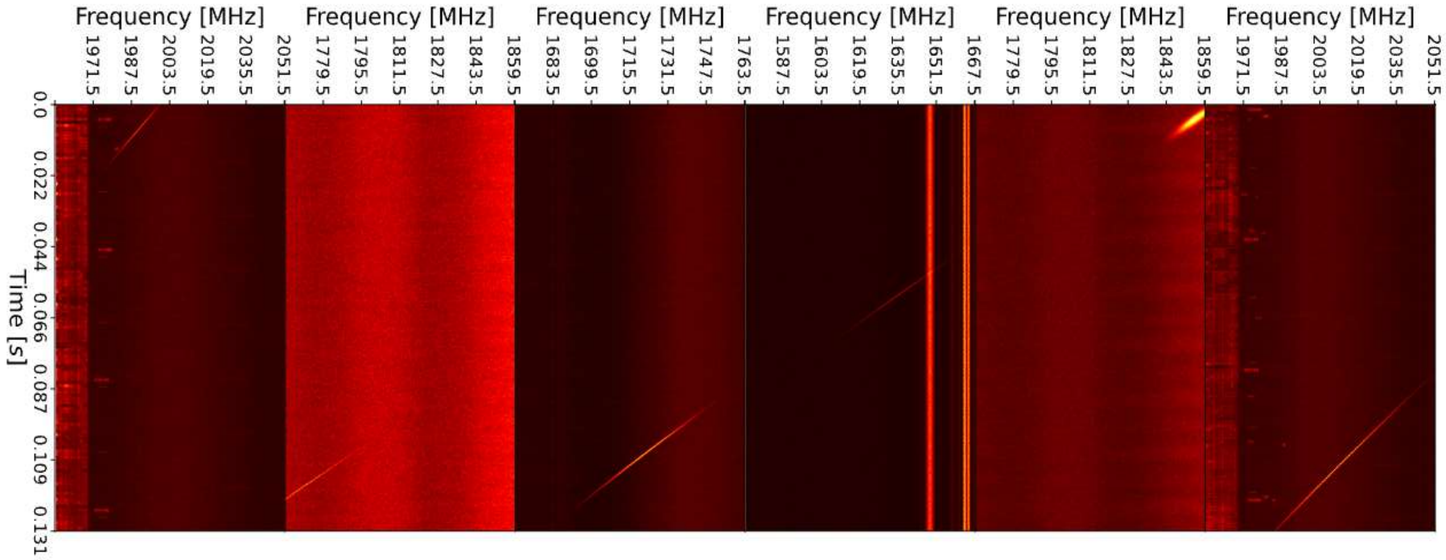
Real-Time End-to-End Deep Learning algorithm for Fast Radio Burst Detection

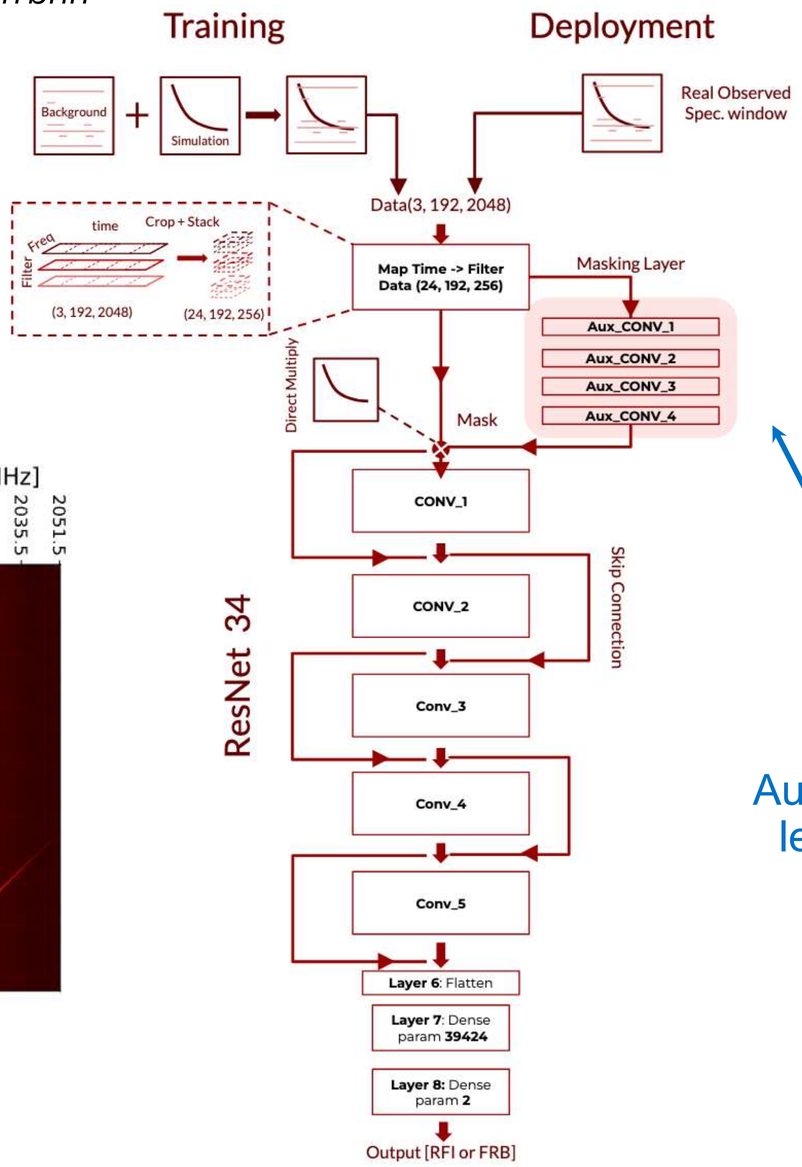
https://github.com/PetchMa/frbnn

Model	Acc. %	Rec. %	FPR %	F1
SPANDAK CNN	92.7	98.0	12.6	0.93
Mask ResNet	99.0	**99.1**	1.18	0.99
8bit Mask ResNet	99.0	98.6	0.60	0.99
Thres. Mask ResNet	**99.1**	98.8	**0.49**	**0.99**
Vanilla ResNet	98.2	98.1	1.78	0.98

Baseline

Our Models





Dedicated Auxiliary Network learns to mask RFI

Enhanced Radio Burst Detection Pipeline

Holoscan Transport

Array Voltages
Circular Buffer

GPUDirect Storage
NVMe Dump

FRBNN
Post-Processing

Expert
TensorRT
Inference
(16+ seconds/hit)

FRBNN
Post-Processing

FRBNN
Post-Processing

BLADE
Beamforming

BLADE
Detection

BLADE
Integration
and Stacking

FRBNN
TensorRT
Inference
(~16 ms/batch)

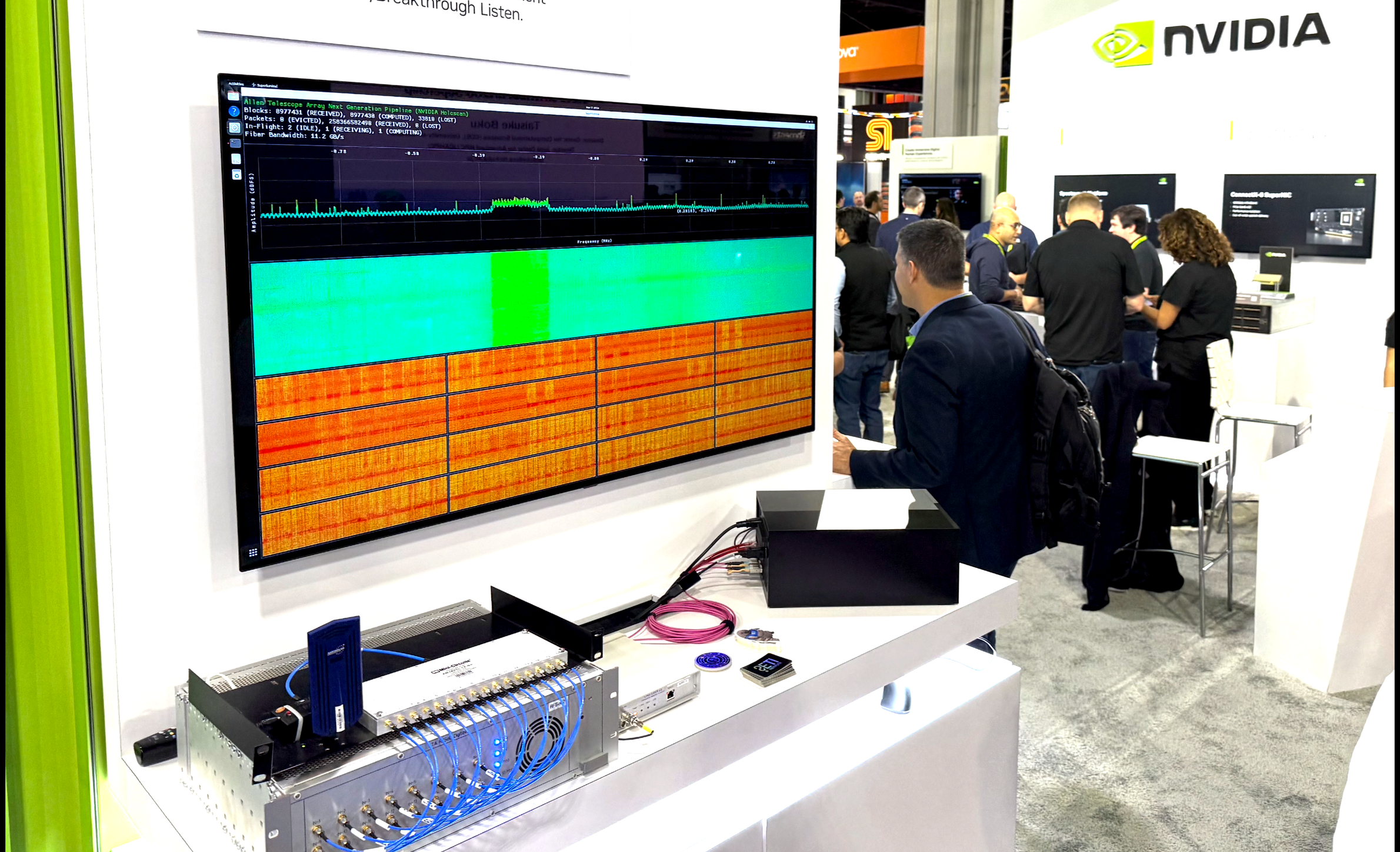
(28, 192, 8192, 2)
86 Gbps

(1, 192, 8192, 2)
12 Gbps

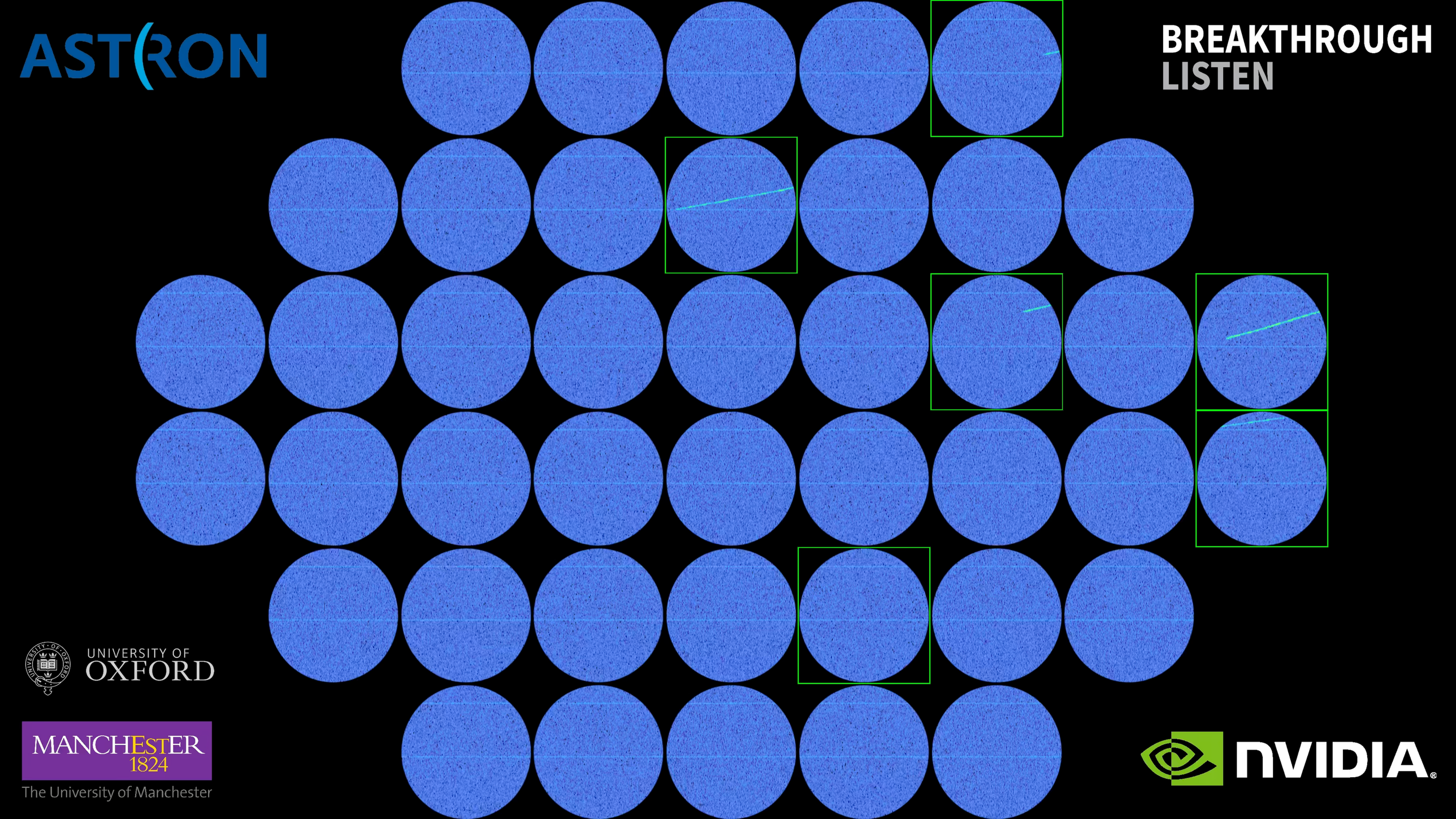
(1, 192, 8192, 1)
6 Gbps

(32, 192, 2048, 1)
~180 Mbps

(A, F, T, P) = A - Antennas, F - Channels, T - Time, P - Polarization


Breakthrough Listen.
Allen Telescope Array Next Generation Pipeline (NVIDIA Holoscan)
Blocks: 8977431 (RECEIVED), 8977430 (COMPUTED), 33818 (LOST)
Packets: 0 (EVICTED), 2583665824928 (RECEIVED), 0 (LOST)
In-Flight: 2 (IDLE), 1 (RECEIVING), 1 (COMPUTING)
Fiber Bandwidth: 11.2 GB/s
Amplitude (dBFS)
Frequency (MHz)
NVIDIA

FUTURE


ASTRON
BREAKTHROUGH
LISTEN
UNIVERSITY OF
OXFORD
MANCHESTER
1824
The University of Manchester
NVIDIA

Next Generation Compute

NVIDIA IGX Orin

- 12-core ARM CPU (Cortex-A78)
- NVIDIA ConnectX-7
  - 2x 100 GbE
  - 32-lane PCIe 5.0 Switch
- NVIDIA A6000 Ada
- OpenBMC (Aspeed AST2600)

Next Generation Compute

Coyote Compute Engine

1x Coyote Machine:

- 4x RTX PRO 6000 Max-Q (384 GB VRAM)
- 4x 200 GbE endpoints (800 Gbps BW)
- 2x PCIe 5.0 NVMe carriers (up to 128 TB)

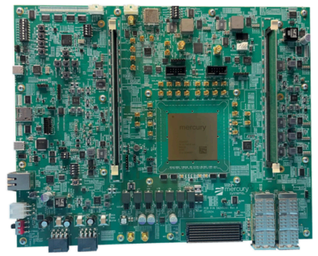
4x Current Pipeline Machines:

- 8x A5000 (192 GB VRAM)
- 8x 100 GbE endpoints (800 Gbps BW)

Next Generation Digitizers

Ultra-Wideband Direct Sampling at the Allen Telescope Array

	Current	*Planned*
RF Bandwidth	~1.5 GHz	~16 GHz
Number of Antennas	28	42
Antenna Data Rate	~46 Gbps	~490 Gbps
Aggregated Data Rate	~1.4 Tbps	~21 Tbps



Mercury

RFS1140 Direct RF SiP

HITEK Systems

Agilex eSOM7C + AD9081

600 gbps/antenna

NVIDIA

Holoscan-based Next Generation Pipeline

Contact:
Luigi Cruz, lfcruz@seti.org

Thank You!

Questions?