Deepwave Digital

Simplifying AI for Communications, Radar, and Wireless Systems

John D. Ferguson

info@deepwavedigital.com

Deep Learning and Radio Frequency (RF) Systems

Deep Learning is Emerging

Cyber

- Intrusion Detection
- Threat classification
- Facial recognition
- Imagery analysis

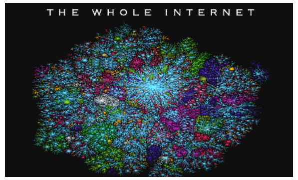
Medicine

- Tumor Detection
- Medical data analysis
- Diagnosis
- Drug discovery

Autonomy

- Pedestrian / obstacle detection
- Navigation
- Street sign reading
- Speech recognition

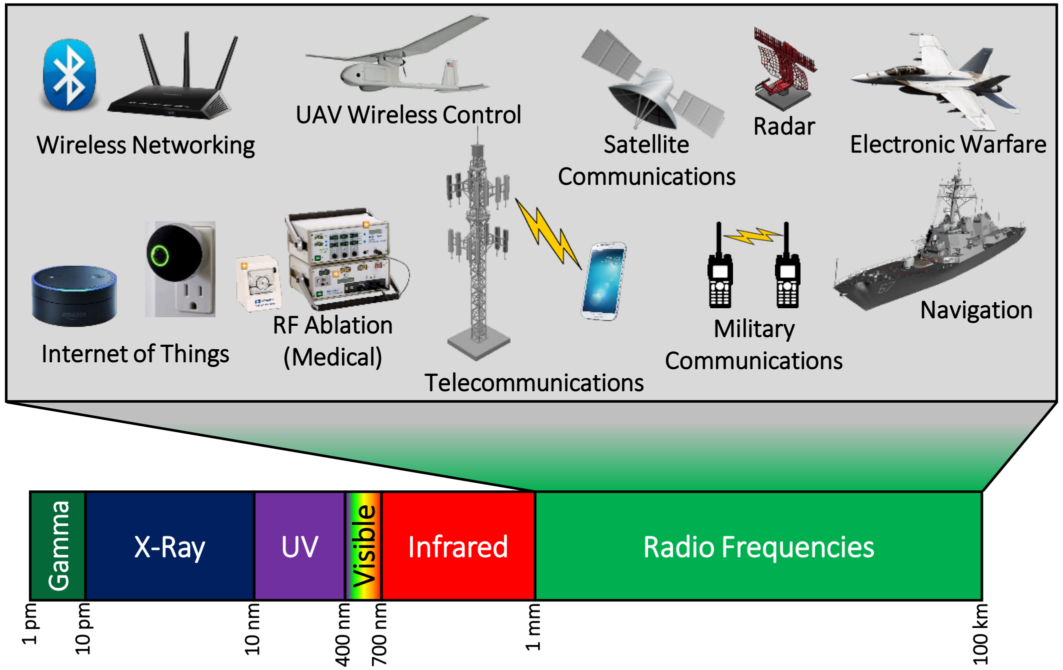
Internet



- Image classification
- Speech recognition
- Language translation
- Document / database searching

Enabled by low-cost, highly capable general purpose graphics processing units (GPUs)

Radio Frequency Technology is Pervasive



Deep learning technology enabled and accelerated by GPU processors

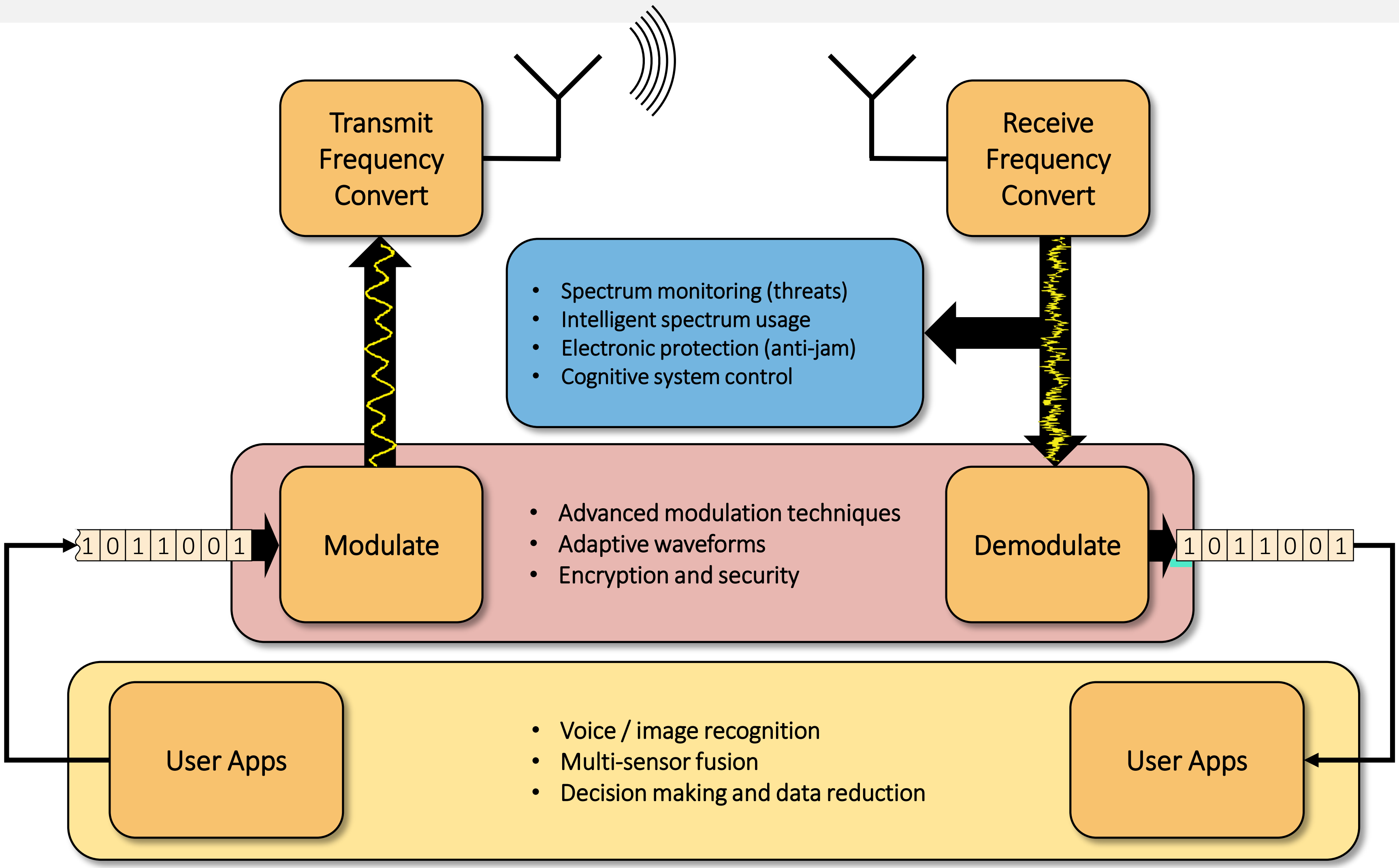
- Has yet to impact design and applications in wireless and radio frequency systems

Where to Use Deep Learning in RF Systems

Spectrum / Network Centric Applications

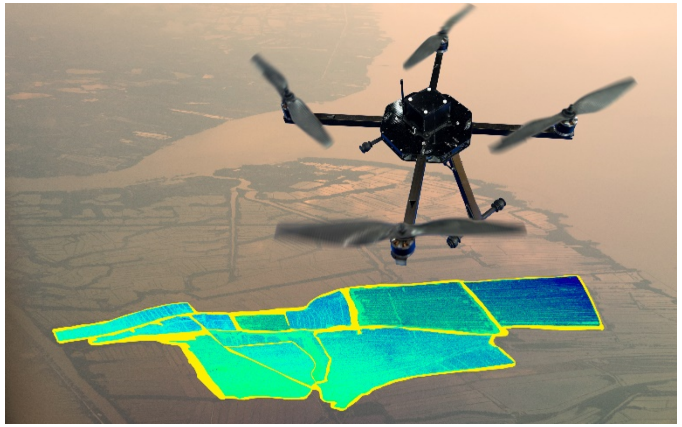
Device / Basestation Centric Applications

User App Centric Applications



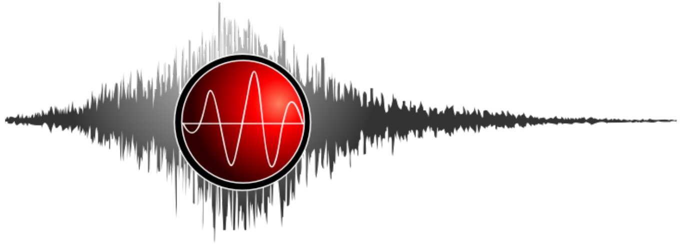
Deep Learning Comparison

Image and Video



- Multiple channels (RGB)
- x, y spatial dependence
- Temporal dependence (video)

Audio and Language



- Single channel
- Frequency, phase, amplitude
- Temporal dependence

Systems and Signals

- Multiple channels
- Frequency, phase, amplitude
- Temporal dependence
- Complex data (I/Q)
- Large Bandwidths
- Human engineered

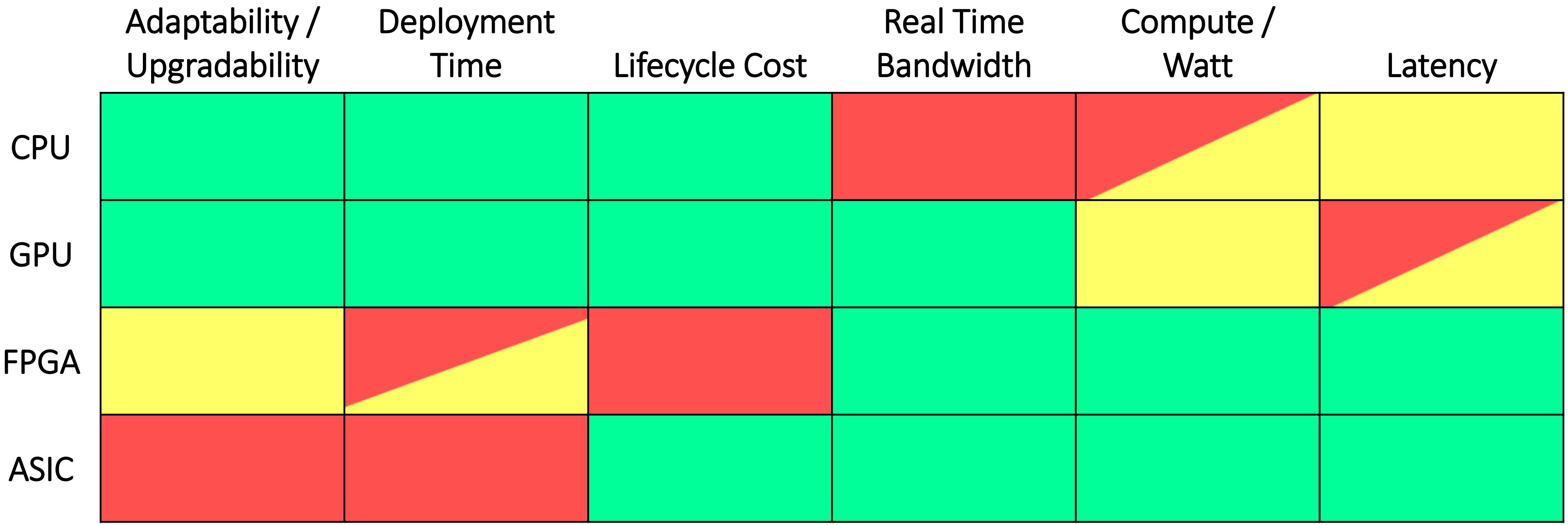
Existing deep learning potentially adaptable to systems and signals

- Must contend with wideband signals and complex data types

Hardware for Deep Learning in RF Systems

	Training		Inference	
	Pros	Cons	Pros	Cons
CPU	• Supported by ML Frameworks • Lower power consumption	• Slower than GPU • Fewer software architectures	• Adaptable architecture • Software programmable • Medium latency	• Low parallelism • Limited real-time bandwidth • Medium power requirements
GPU	• Supported by ML Frameworks • Widely utilized • Highly parallel / adaptable • Good throughput vs power	• Overall power consumption • Requires highly parallel algorithms	• Adaptable architecture • High real-time bandwidth • Software programmable	• Medium power requirements • Not well integrated into RF • Higher latency
FPGA	Not widely utilized, not well suited (yet)		• High power efficiency • High real-time bandwidth • Low latency	• Long development / upgrades • Limited reprogrammability • Requires special expertise
ASIC	Not widely utilized, not well suited		• Extremely power efficient • High real-time bandwidth • Highly reliable • Low latency	• Extremely expensive • Long development time • No reprogrammability • Requires special expertise

Critical Performance Parameters for Deep Learning in RF Systems



GPU signal processing can provide wideband capability and software upgradability at lower cost and development time

- Must contend with increased latency (~2 microsecond)

Outline

- Introduction to Deep Learning in RF
- Deepwave's Technology
- Signal Detection and Classification
- Real-time Benchmarks on Embedded GPUs
- Summary

Why Has Deep Learning in RF Not Been Addressed

Bandwidth Limitations

remote processing not possible

Limited Compute Resources

at field site

Complicated Software

for RF and AI independently

- AI requires large data sets
- Insufficient bandwidth to send to remote data center

- No RF systems exist with integrated AI computational processors

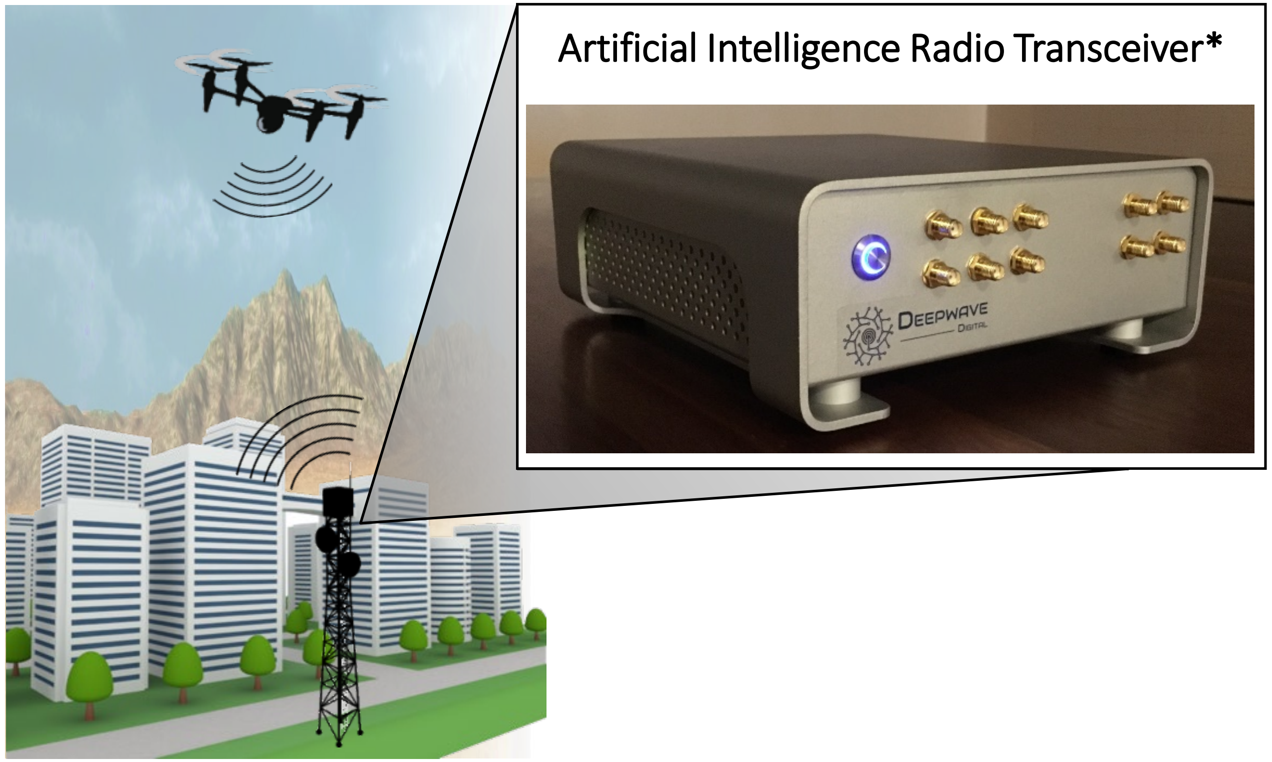
- Disjointed software
- Difficult to program and understand

Deepwave's Software Defined Radio

A Platform for a Multitude of Applications

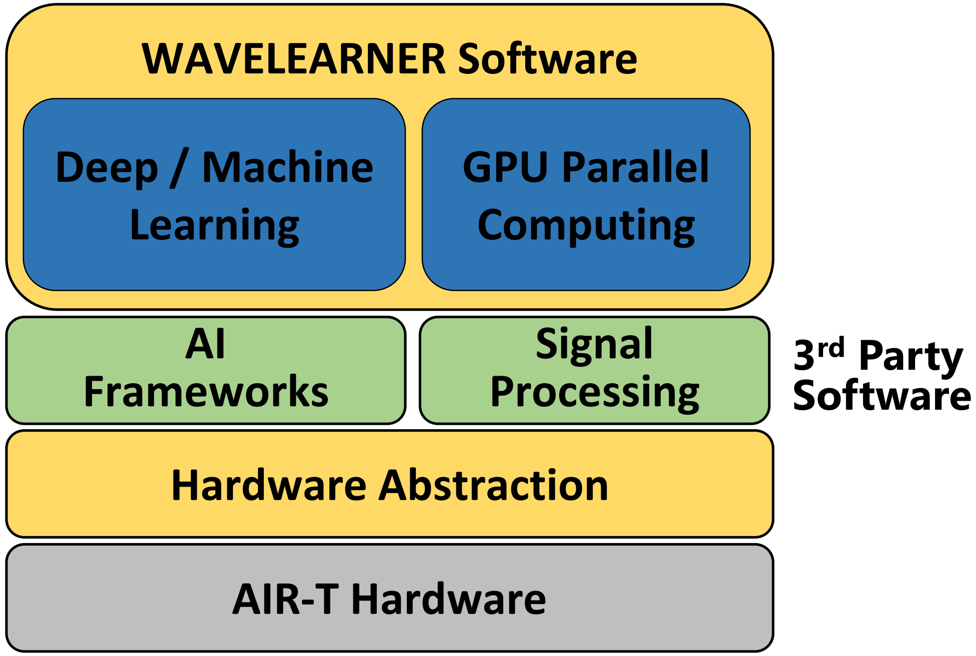
The Platform

Complete Edge-compute AI Platform for RF



The Software

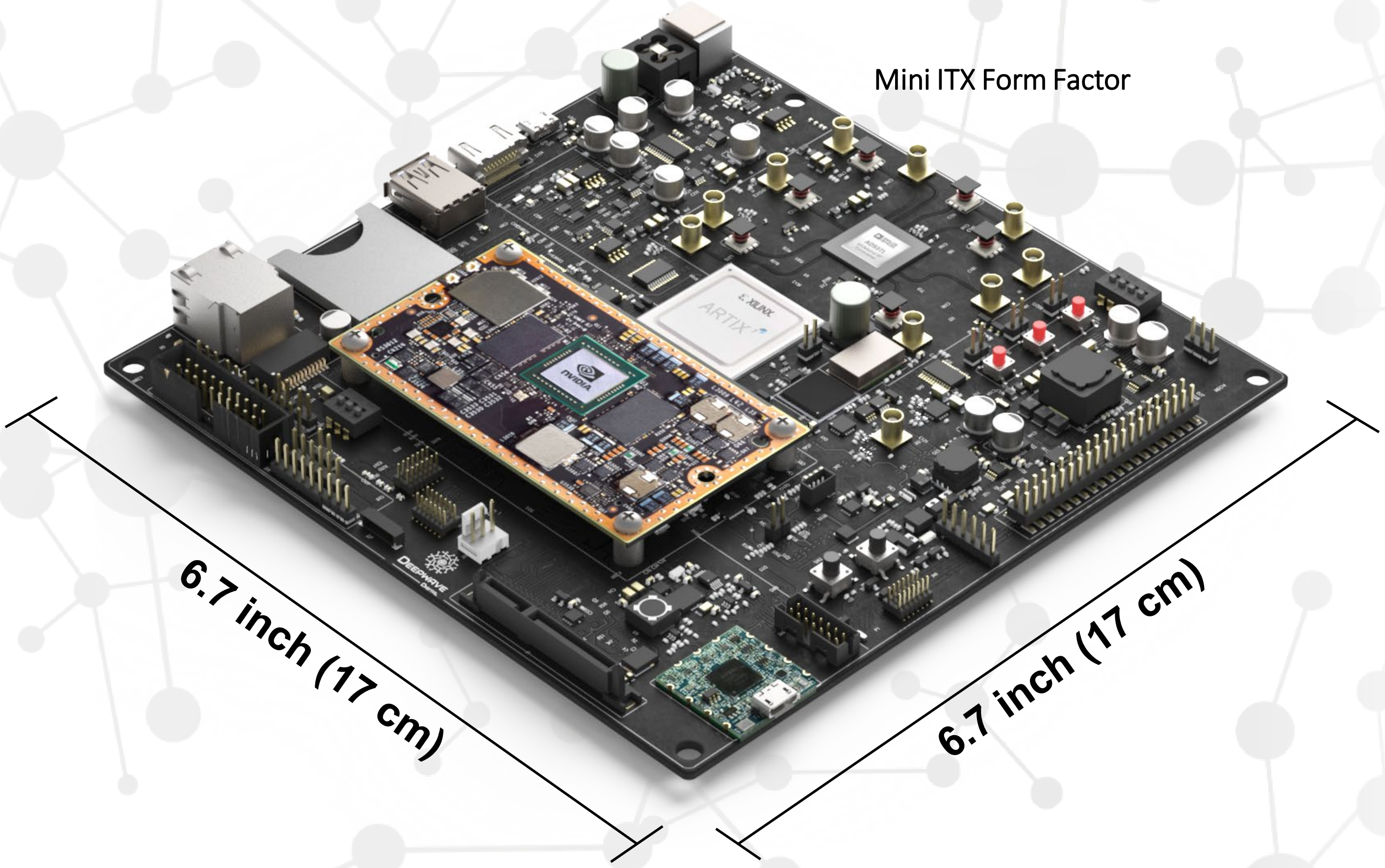
Simply build AI into wireless technology



*Patent Pending

Artificial Intelligence Radio Transceiver (AIR-T)

AIR-T

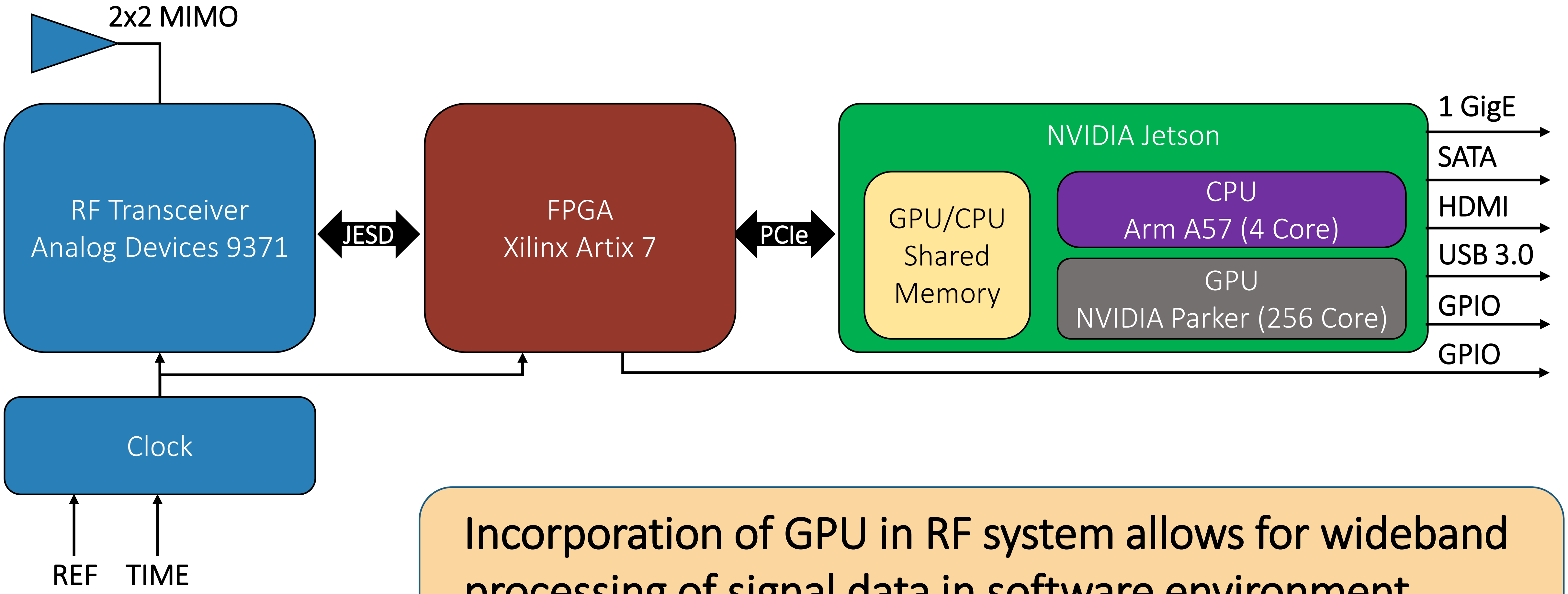


Hardware Specifications

- 2x2 MIMO Transceiver
  - Analog Devices 9371 chip
  - Tunable from 300 MHz to 6 GHz
  - 100 MHz bandwidth per channel
- Digital Signal / Deep Learning Processors
  - Xilinx Artix 7 FPGA
  - NVIDIA Jetson TX2
    - ARM Cortex-A57 (quad-core)
    - Denver2 (dual core)
    - Nvidia Pascal 256 Core GPU
    - Shared GPU/CPU memory
- Connectivity
  - 1 PPS / 10 MHz for GPS Synchronization
  - External LO input
  - HDMI, USB 2.0/3.0, SATA, Ethernet, SD Card, GPIO

Artificial Intelligence Radio Transceiver (AIR-T)

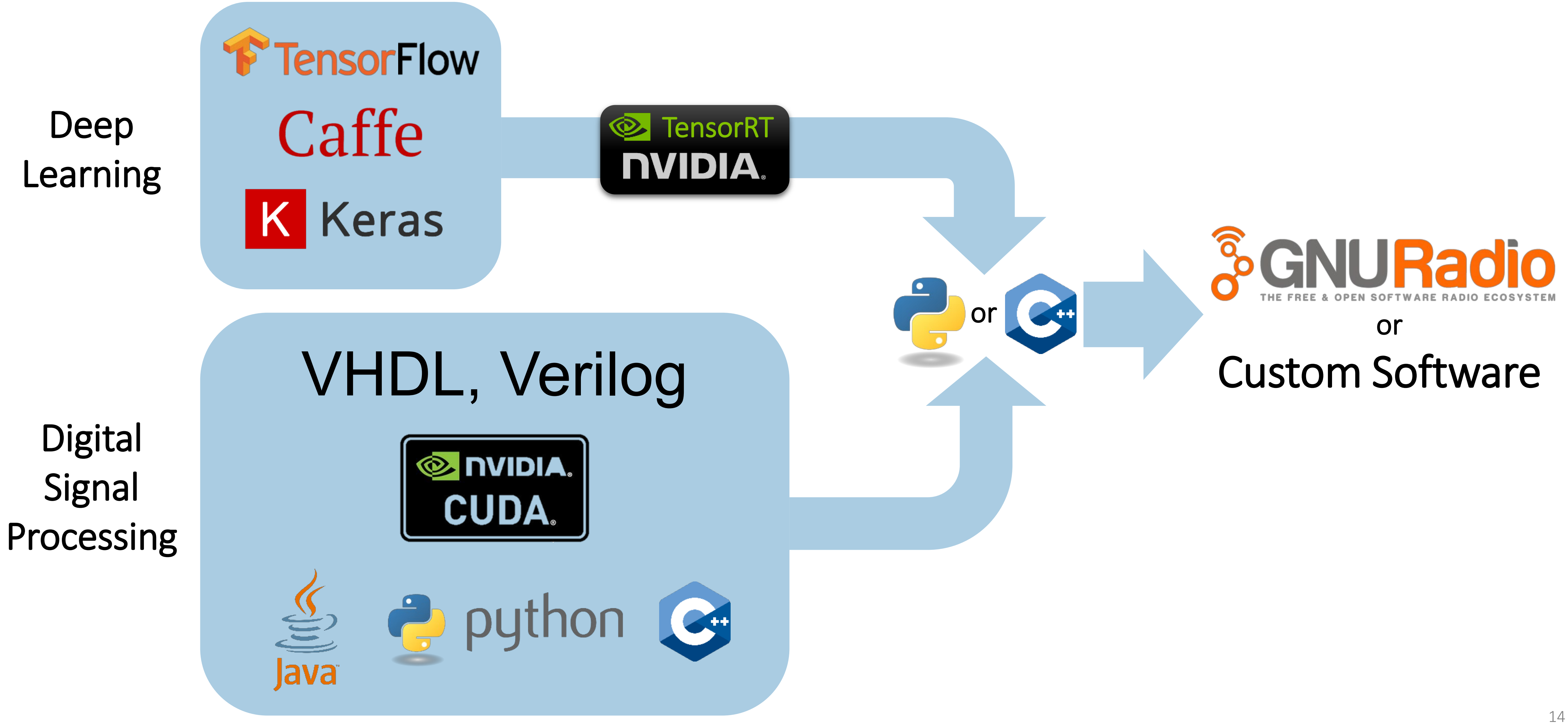
Block Diagram



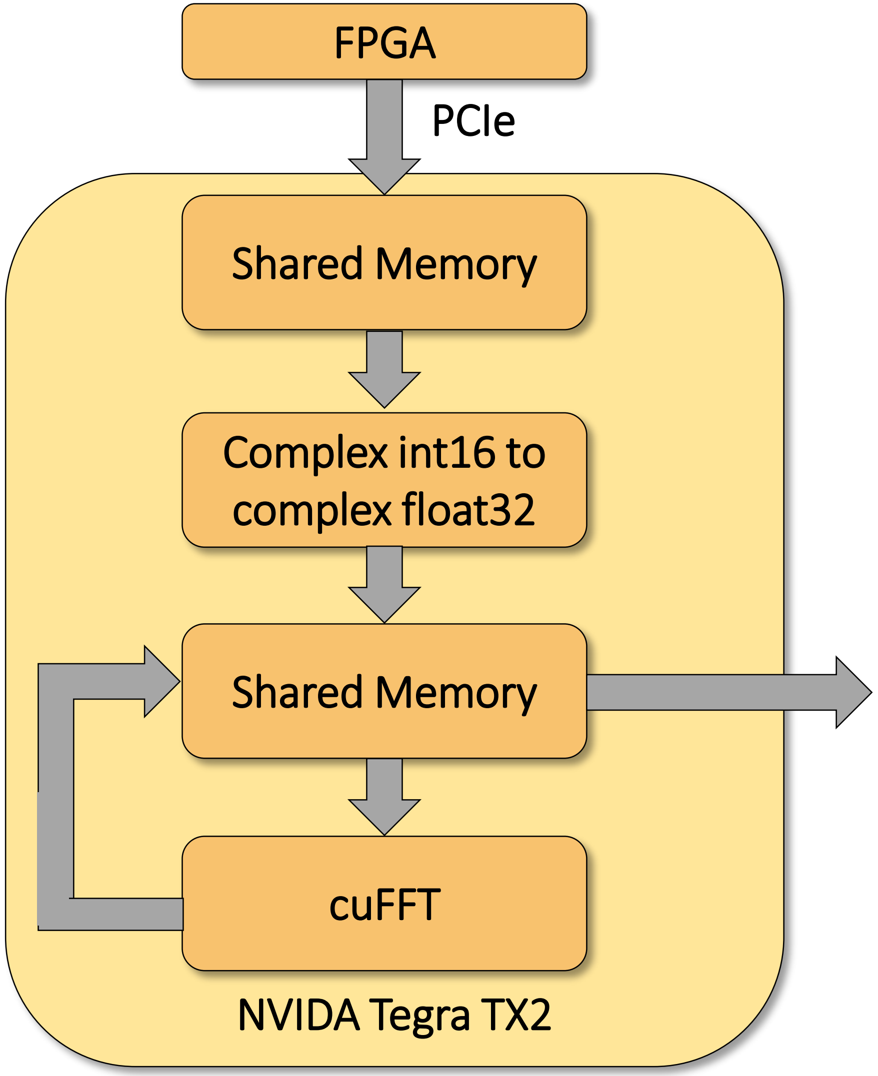
Incorporation of GPU in RF system allows for wideband processing of signal data in software environment

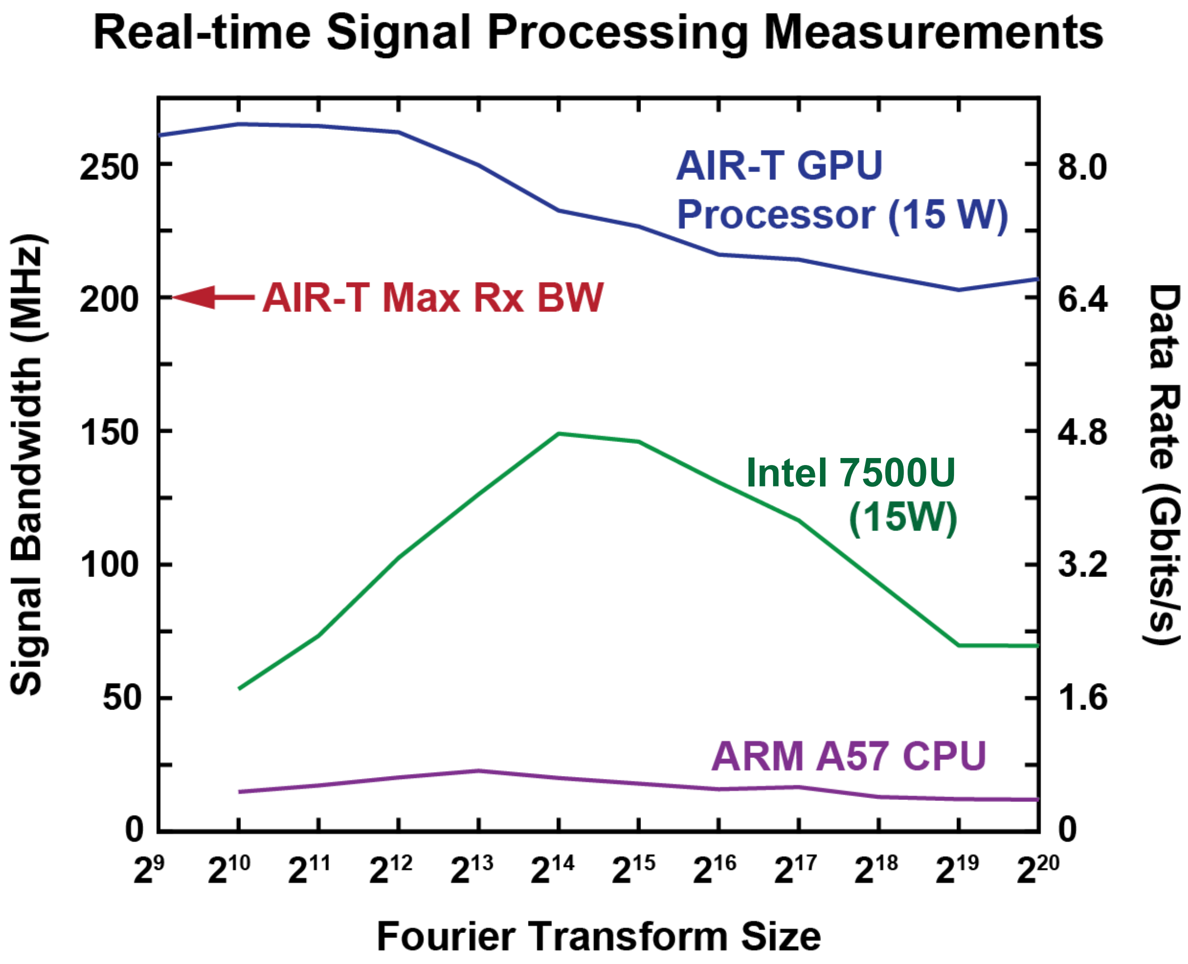
- Reduces development time and cost

Simplified Programming

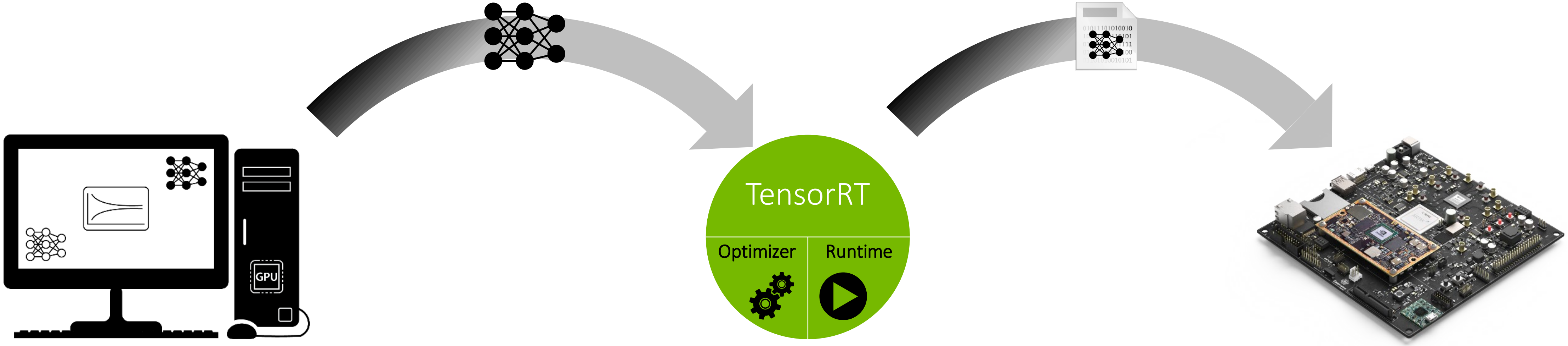


FFT Performance Testing





Inference at the Edge with GR-Wavelearner



Train Neural Network

Optimize Neural Network

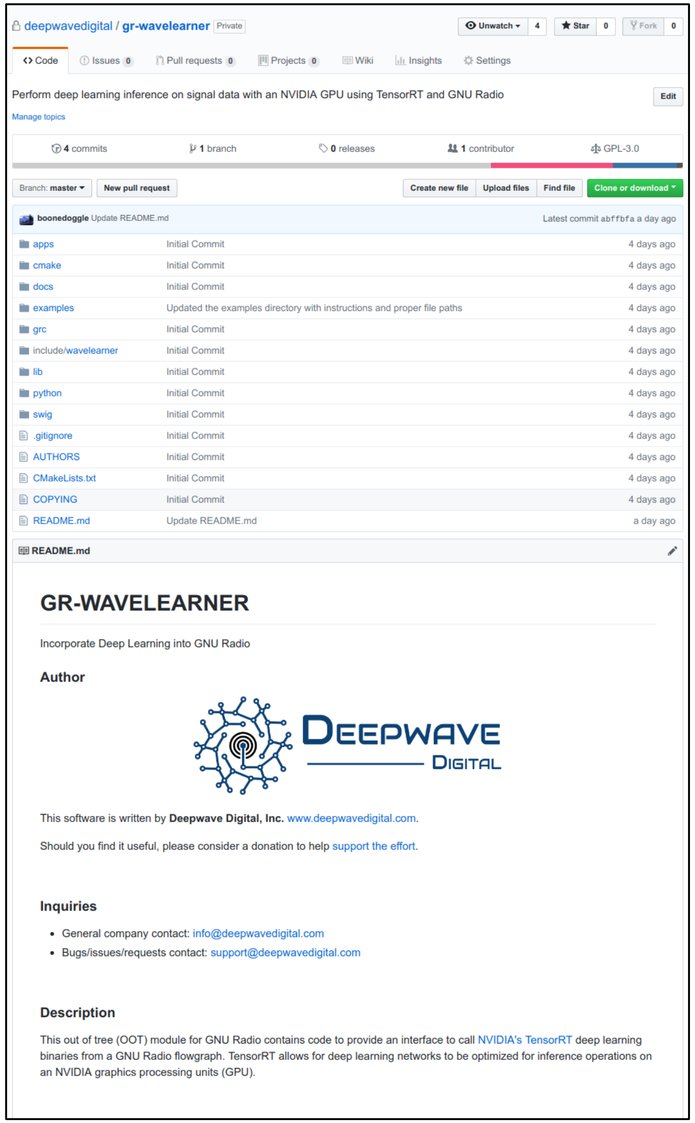
Deploy Application

TensorFlow
PyTorch
mxnet
Cognitive Toolkit
Caffe2
PaddlePaddle
K
Chainer

TensorRT
NVIDIA

DEEPWAVE DIGITAL
GNURadio

GR-Wavelearner Software

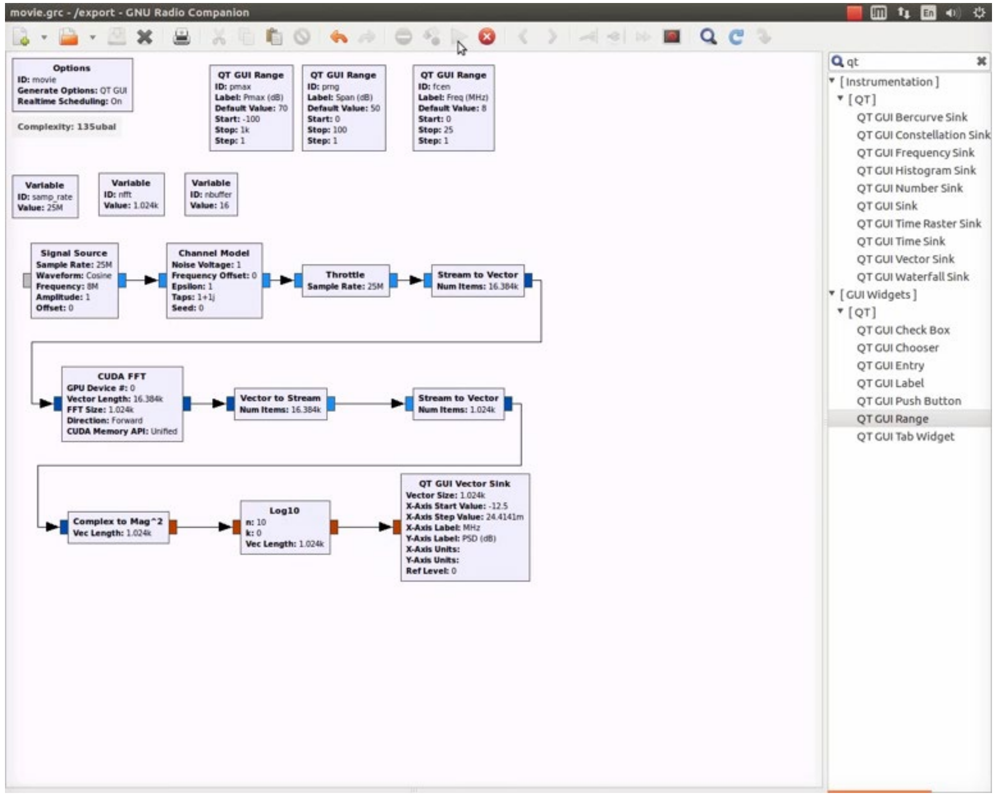


- Goal is to help the open source community easily deploy deep learning within signal processing applications
- Well documented README with dependency installation instructions to get started quickly
  - Ubuntu 16.04 recommended, Windows 10 supported
  - NVIDIA Docker Container 18.08*
- Signal classifier example provided:
  - GNU Radio Flowgraph
  - Python source code
  - PLAN files that are executable on the AIR-T and Maxwell
  - Signal data file example for testing
- Support for TensorRT 5.0
- Available at: deepwavedigital.com/wavelearner

https://docs.nvidia.com/deeplearning/sdk/tensorrt-container-release-notes/rel_18.08.html

GNU Radio – Software Defined Radio (SDR) Framework

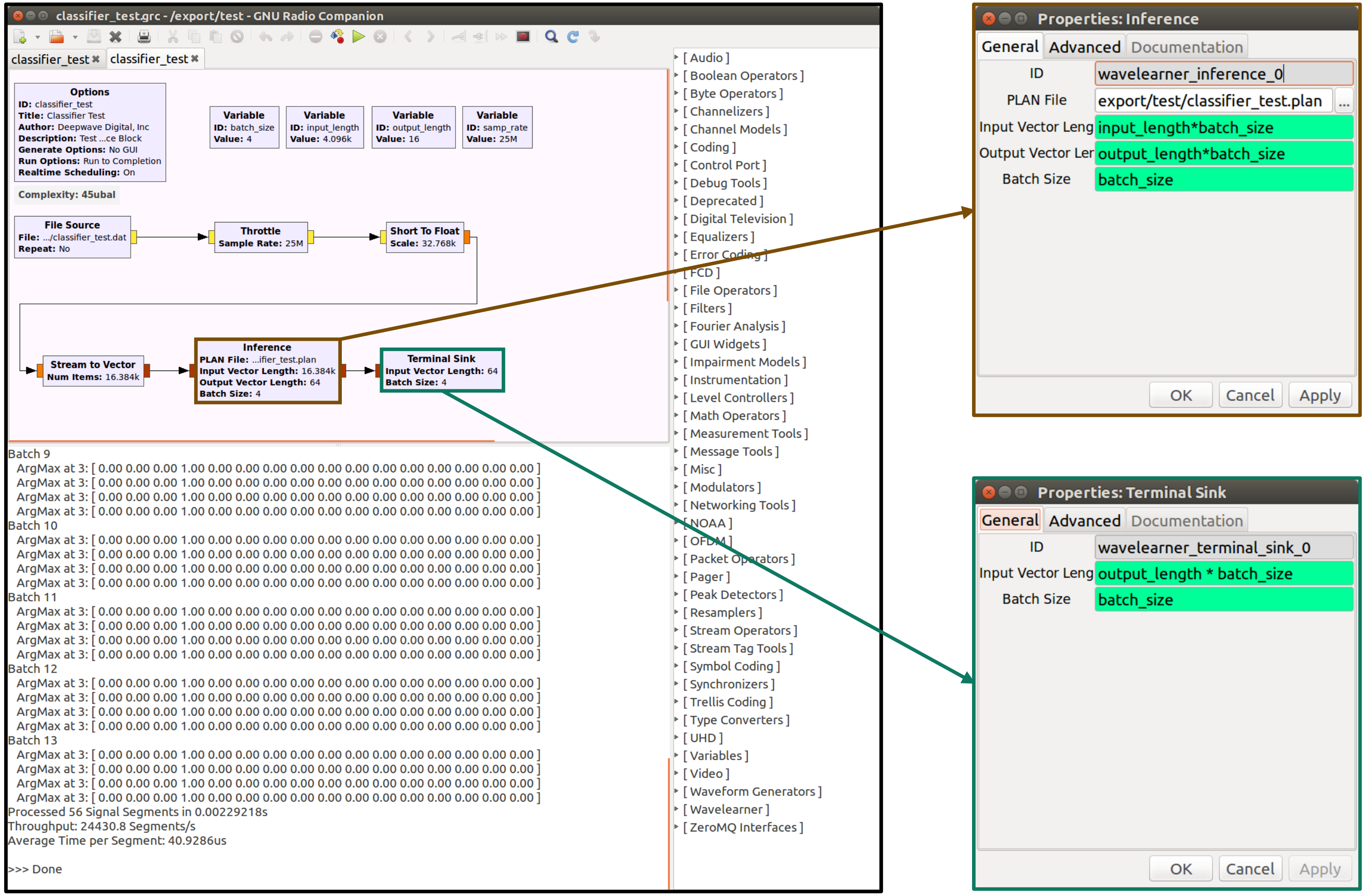
- Popular open source software defined radio (SDR) toolkit:
  - RF Hardware optional
  - Can run full software simulations
- Python API
  - C++ under the hood
- Easily create DSP algorithms
  - Custom user blocks
- Primarily uses CPU
  - Advanced parallel instructions
  - Recent development: RFNoC for FPGA processing
- Deepwave is integrating GPU support for both DSP and ML



GNURadio
THE FREE & OPEN SOFTWARE RADIO ECOSYSTEM

GR-Wavelearner

- Out of tree (OOT) module for GNU Radio
- Allows users to easily incorporate deep learning into signal processing
- C++ and Python API
- Open source GPLv3 license
- Two blocks currently:
  - Inference – TensorRT wrapper for GNU Radio
  - Terminal Sink – Python module for displaying classifier output



Outline

- Introduction to Deep Learning in RF
- Deepwave's Technology
- Signal Detection and Classification
- Real-time Benchmarks on Embedded GPUs
- Summary

Multi-transmitter Environmental Scenario

AIR-T monitors congested spectrum using deep learning

Radar Signal Detector Model: Transmitted Signals

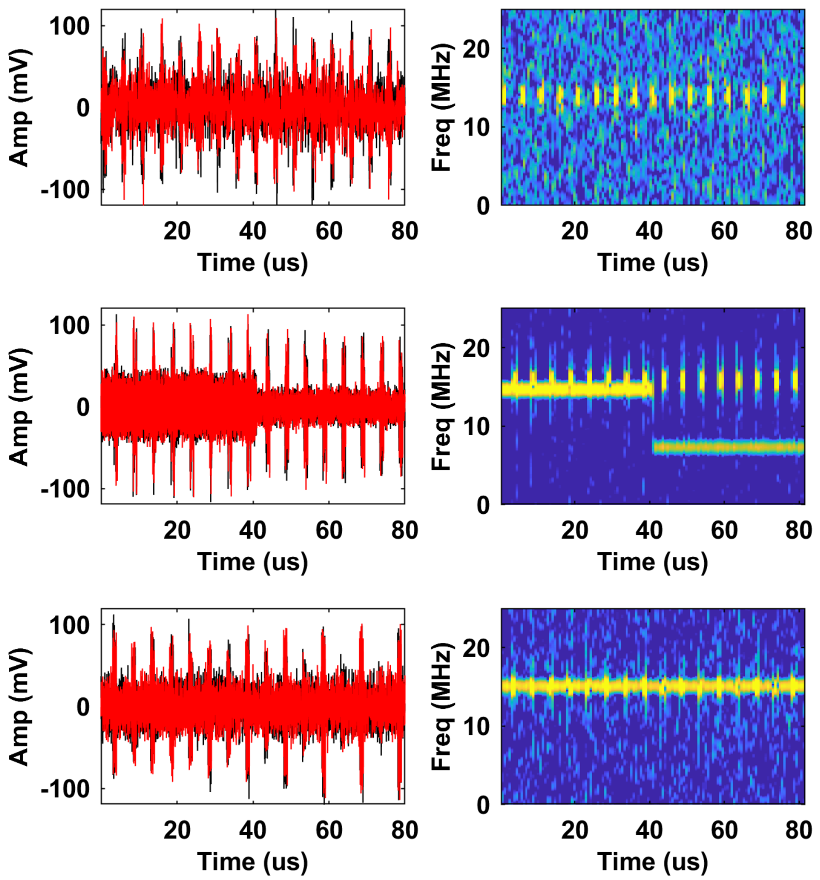
Radar Waveform	Nothing	Interference	Surveillance	Ground (LFM1)	Ground (LFM2)	MTI	Airborne (Med PRF)	Airborne (High PRF)	Ground (Frank Code)	Nautical (Short Range)	Nautical (Long Range)	Nautical (Long Range)	Ground (NLFM1)	Ground (NLFM2)	Ground (NLFM3)
Linear Pulse			X	X	X					X	X	X			
Non-Linear Pulse													X	X	X
Phase Coded Pulse									X						
Pulsed Doppler						X	X	X							

Technique demonstration shown with nominal radar signals

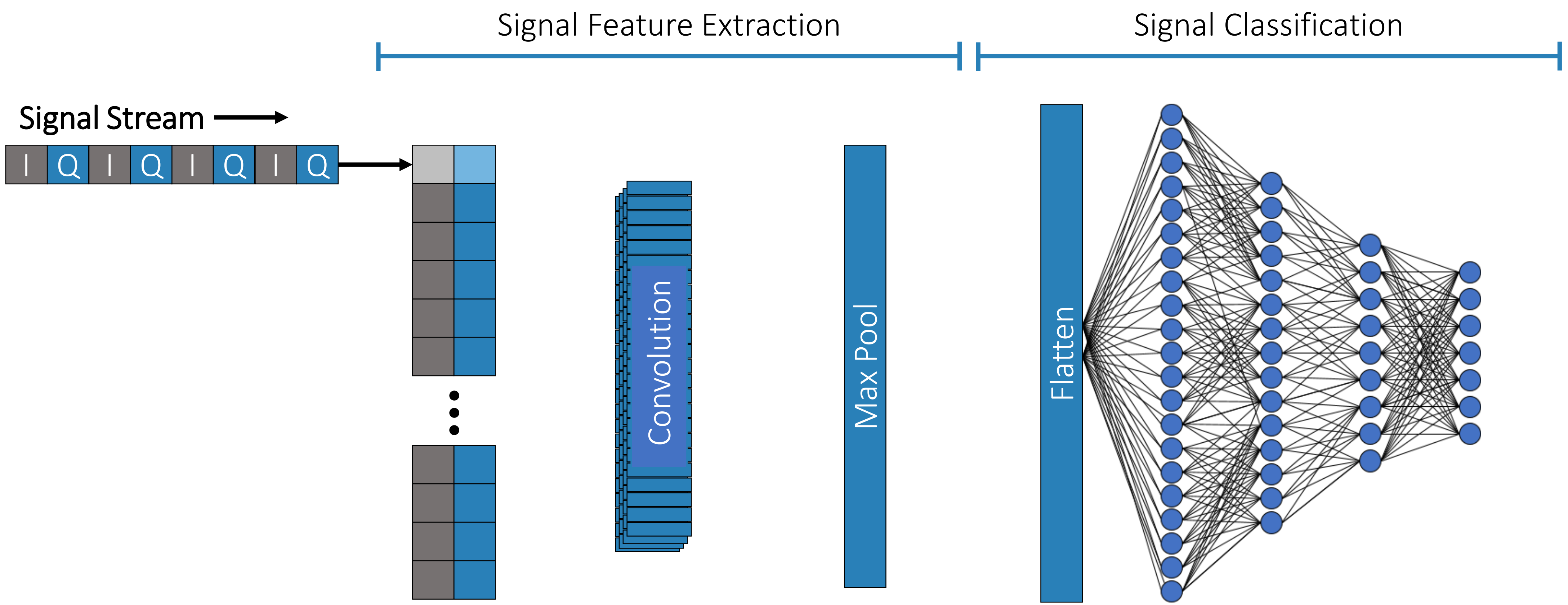
- Method applicable to communications, cellular, and other RF protocols

Dataset Overview

- Goal: Develop a deep learning classifier that detects signals below noise floor
  - Requires training on noisy data with and without interference
- Swept SNIR from -35 dB to 20 dB in 1 dB increments
  - 1000 training segments per SNIR
  - 500 inference segments per SNIR
  - Up to 3 interferers in each segment

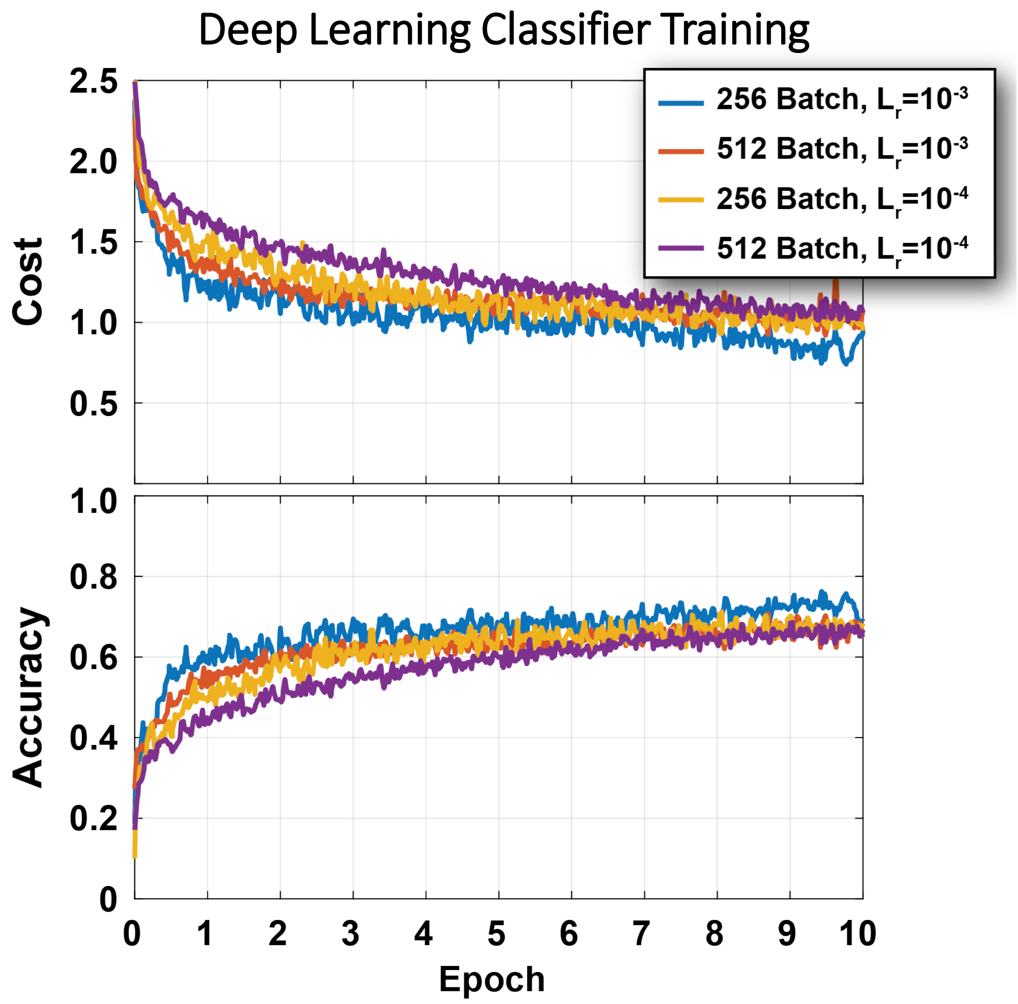


Radar Signal Detector Model: Example Classifier

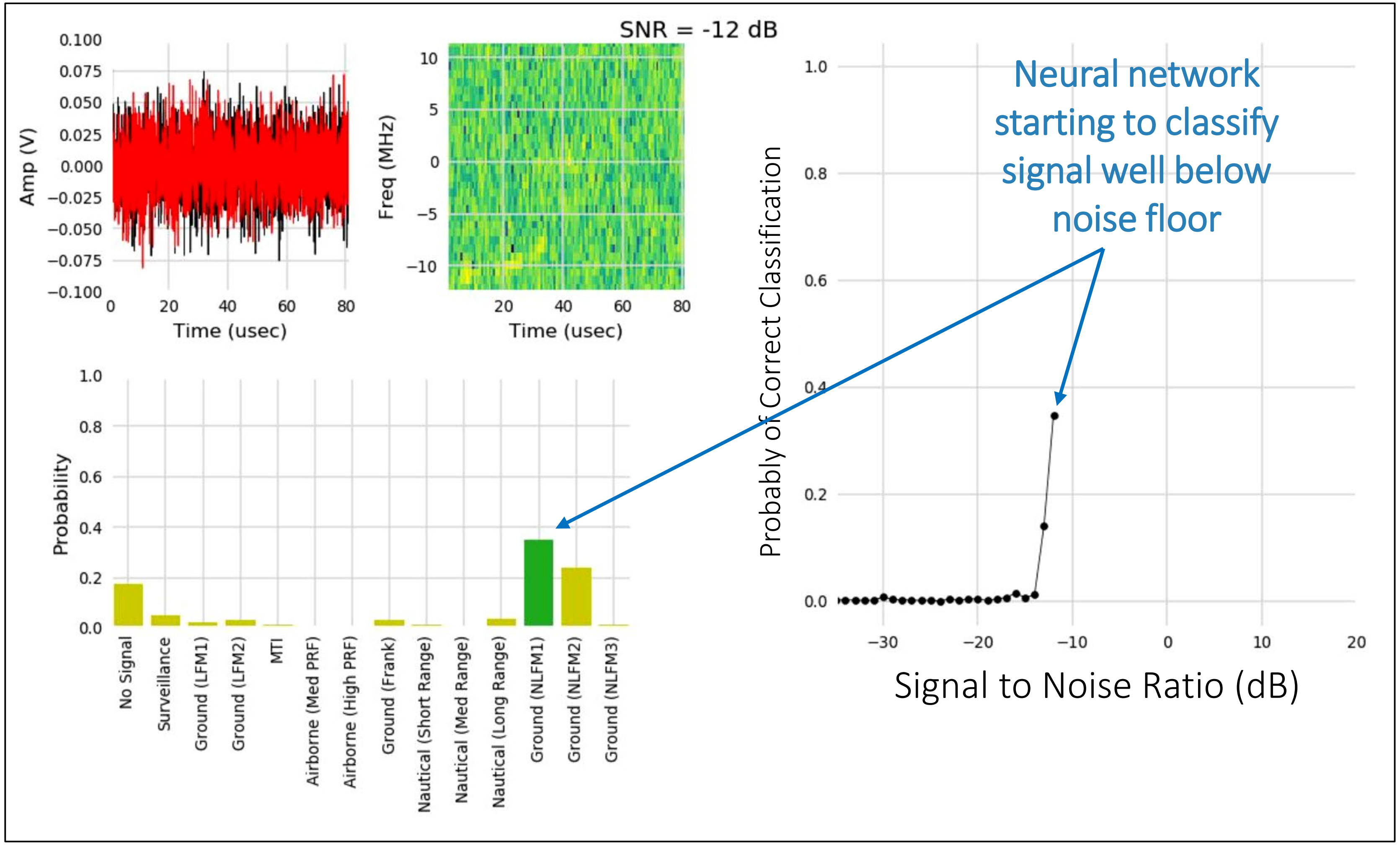


Training Process and Progress

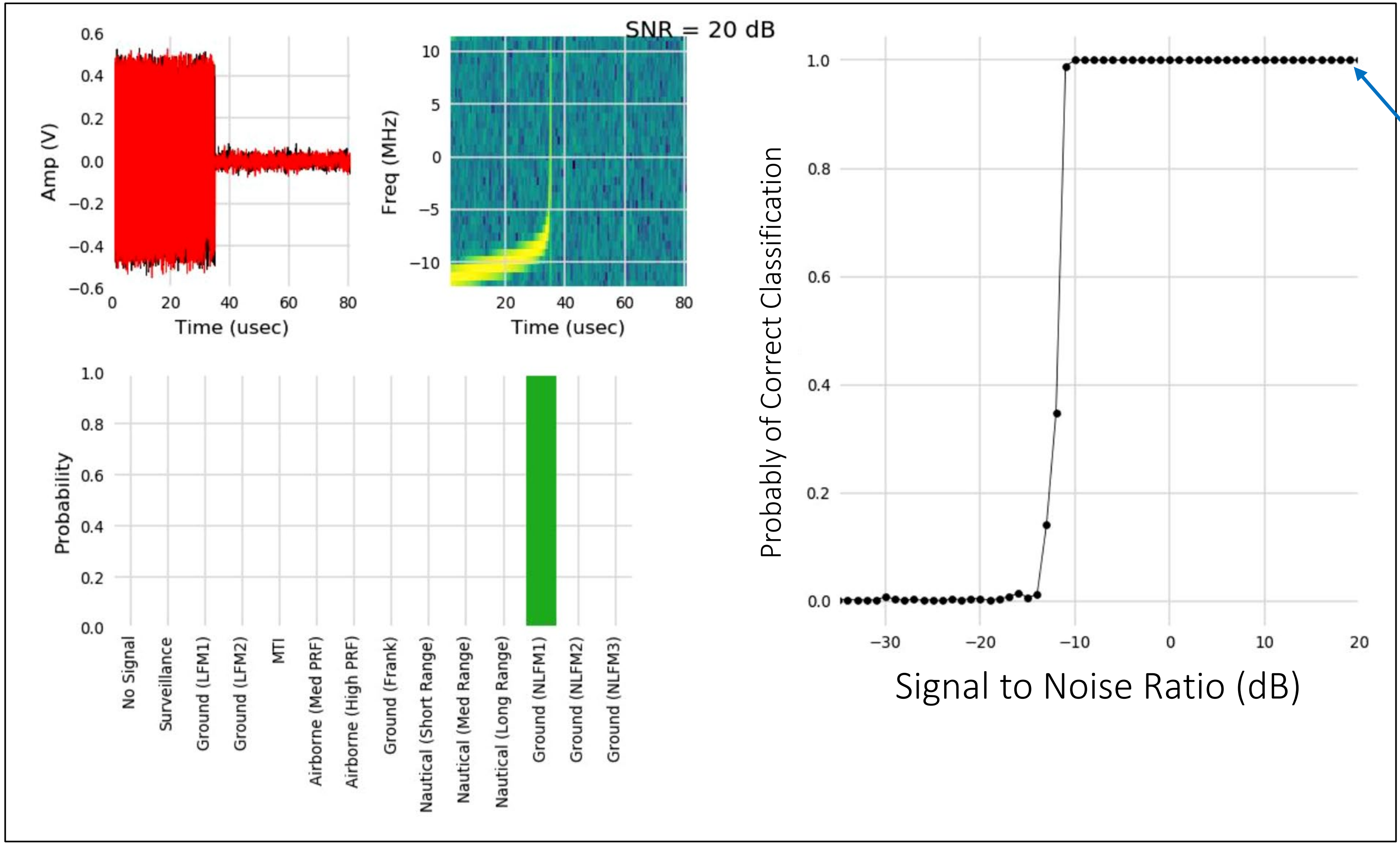
- 1000 training segments per SNR
  - 55 different SNR values
- Training on low SNR values increase detection sensitivity
- 100% accuracy not expected due to training at extremely low SNR values
- Softmax cross entropy
- Adam Optimizer



Detecting and Classifying Low Power Signals



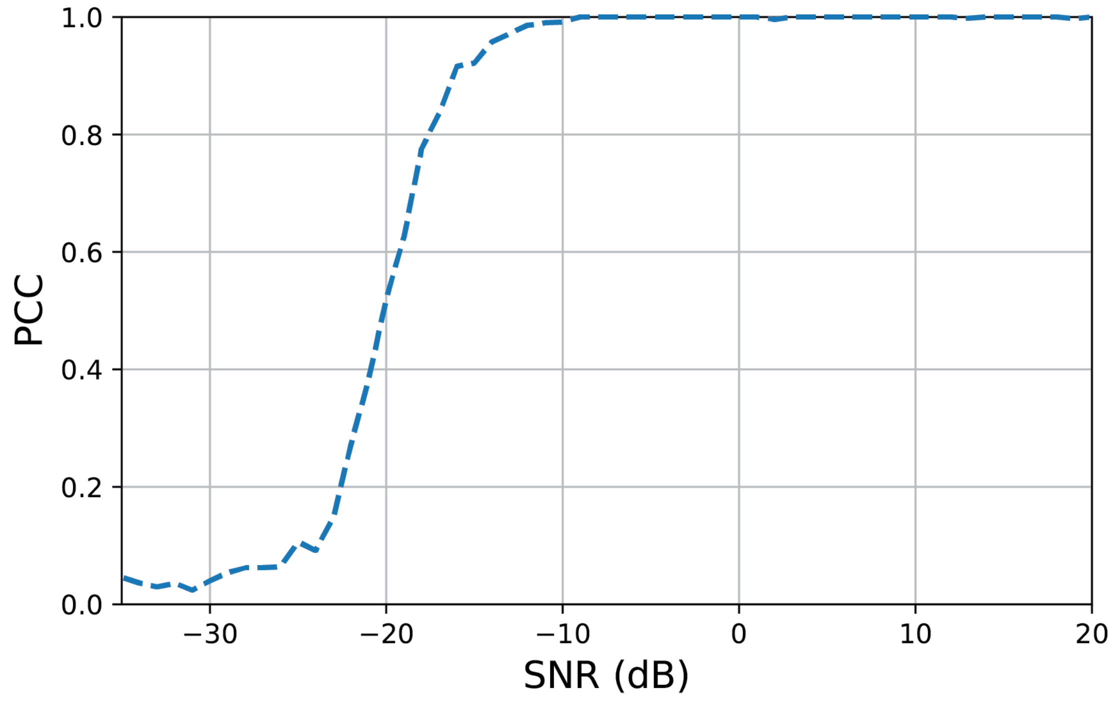
Detecting and Classifying Low Power Signals



Near 100% classification probability with SNR > -10dB

Receiver Operating Characteristic (ROC) Curve

Probability of Correct Classification for One Radars

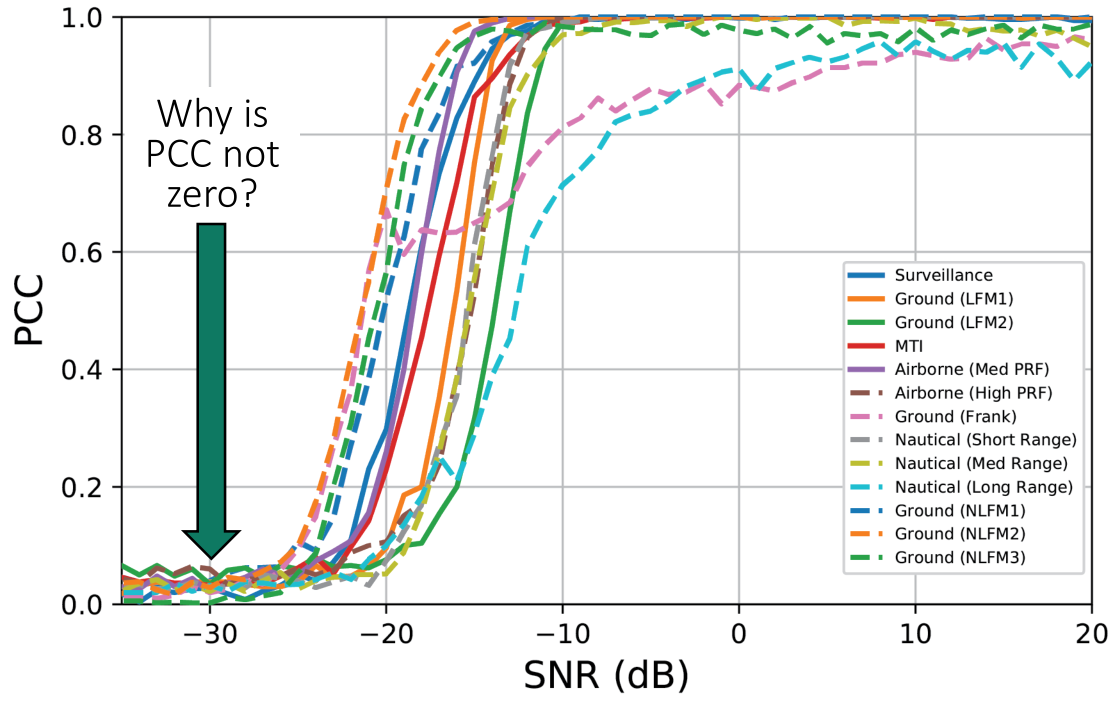


Decibel (dB) Refresher

Signal-to-Noise Ratio (dB)	Receiver Noise Power (milliwatts)	Received Signal Power (milliwatts)
20	1	100
10	1	10
0	1	1
-10	1	0.1
-20	1	0.01
-30	1	0.001

Receiver Operating Characteristic (ROC) Curve

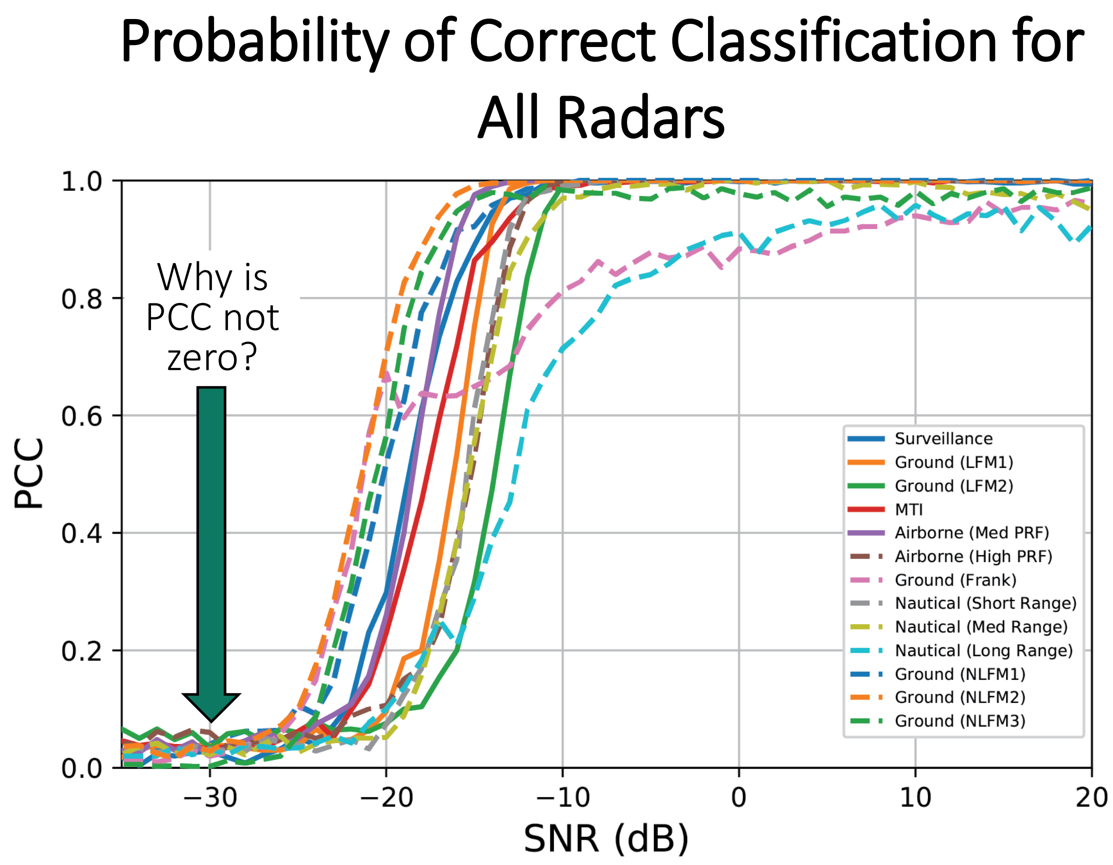
Probability of Correct Classification for All Radars

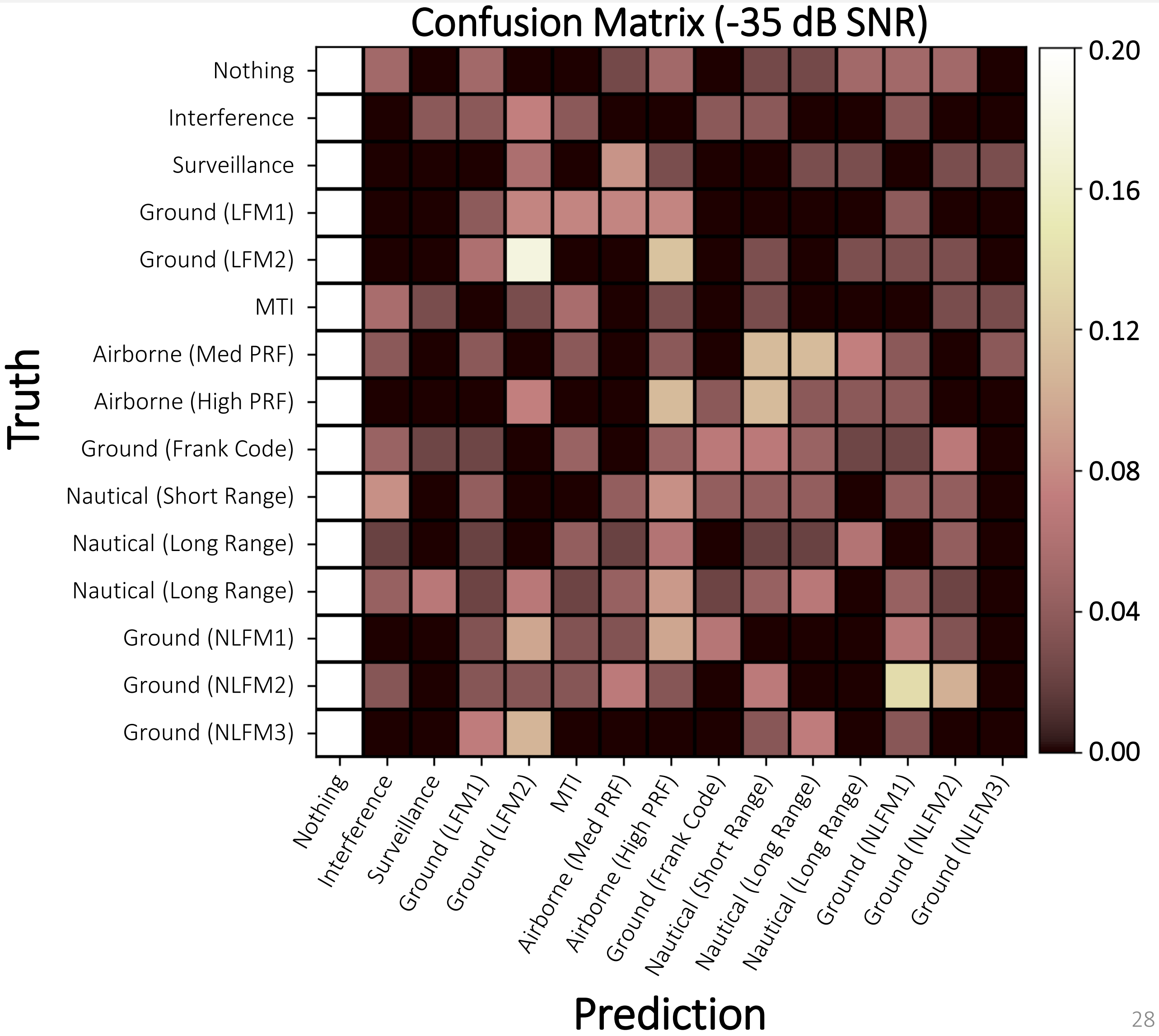


Decibel (dB) Refresher

Signal-to-Noise Ratio (dB)	Receiver Noise Power (milliwatts)	Received Signal Power (milliwatts)
20	1	100
10	1	10
0	1	1
-10	1	0.1
-20	1	0.01
-30	1	0.001

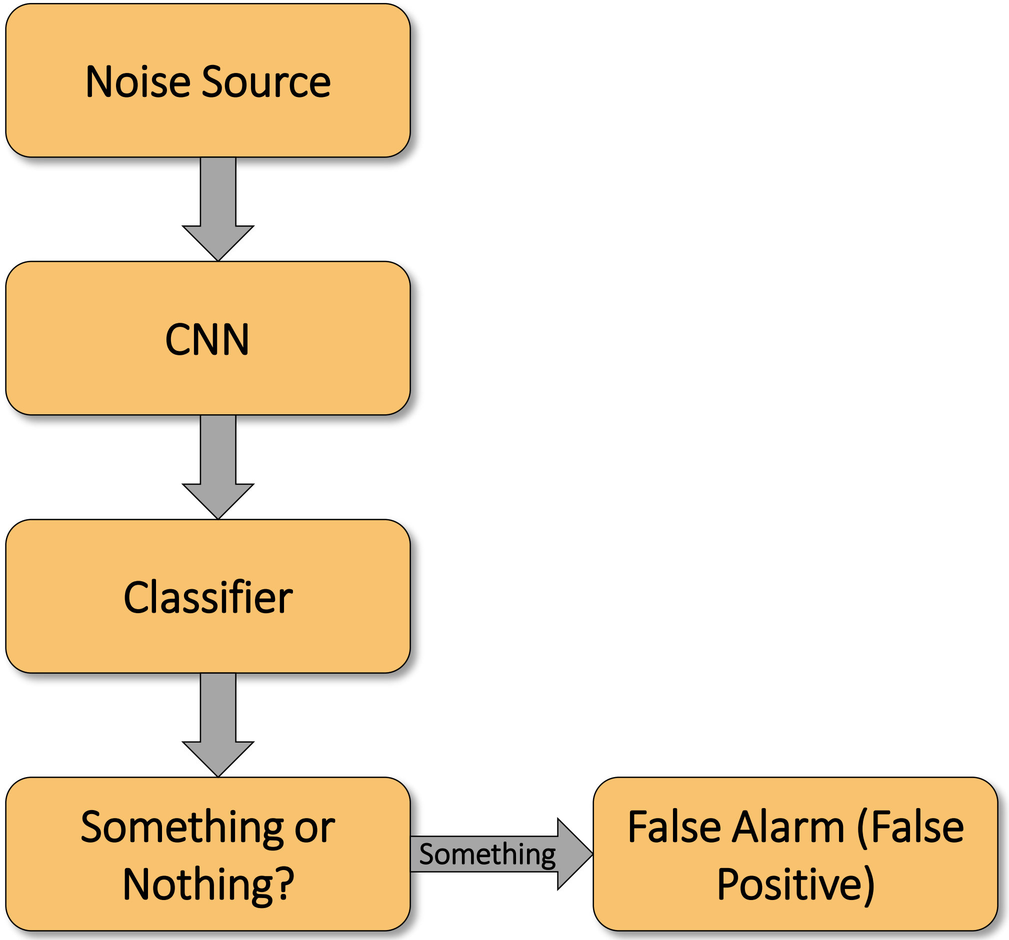
Receiver Operating Characteristic (ROC) Curve

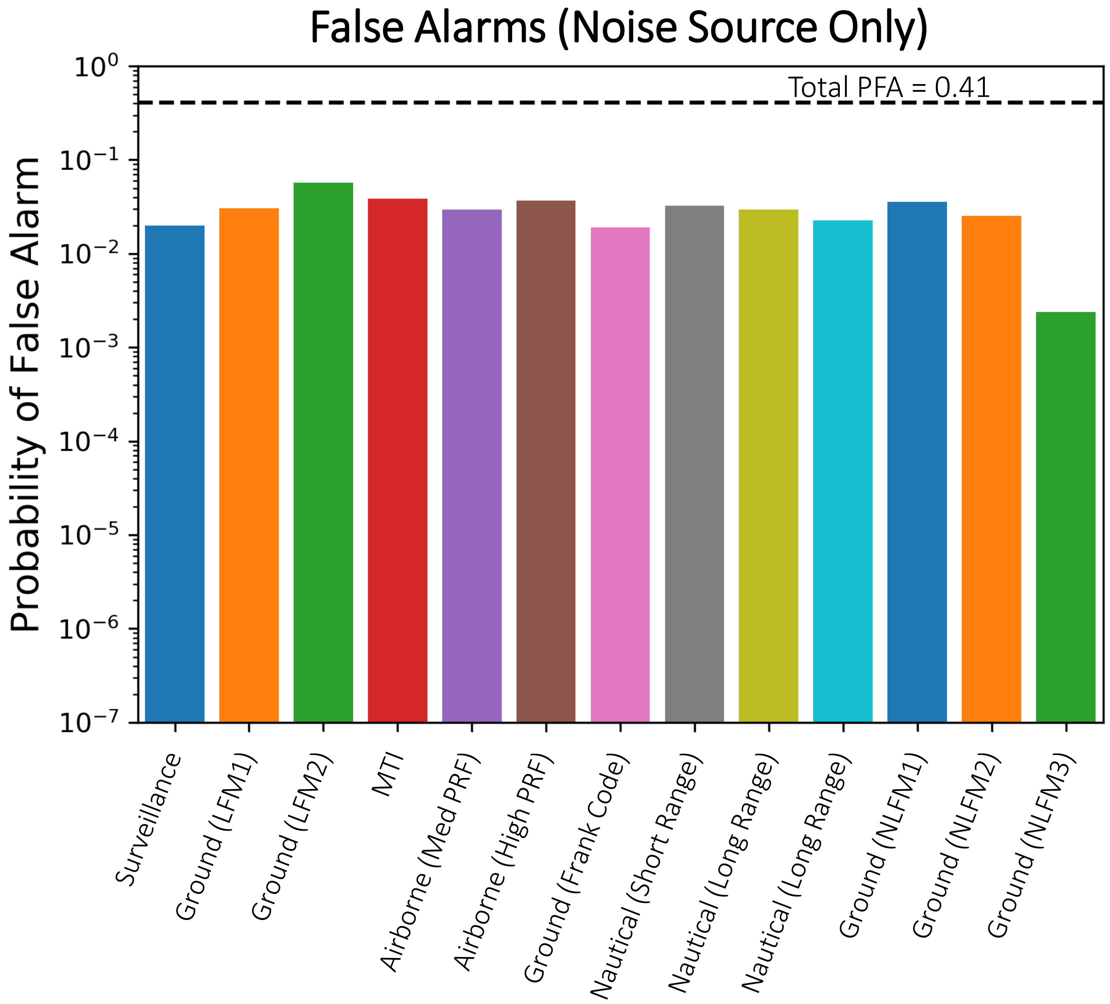




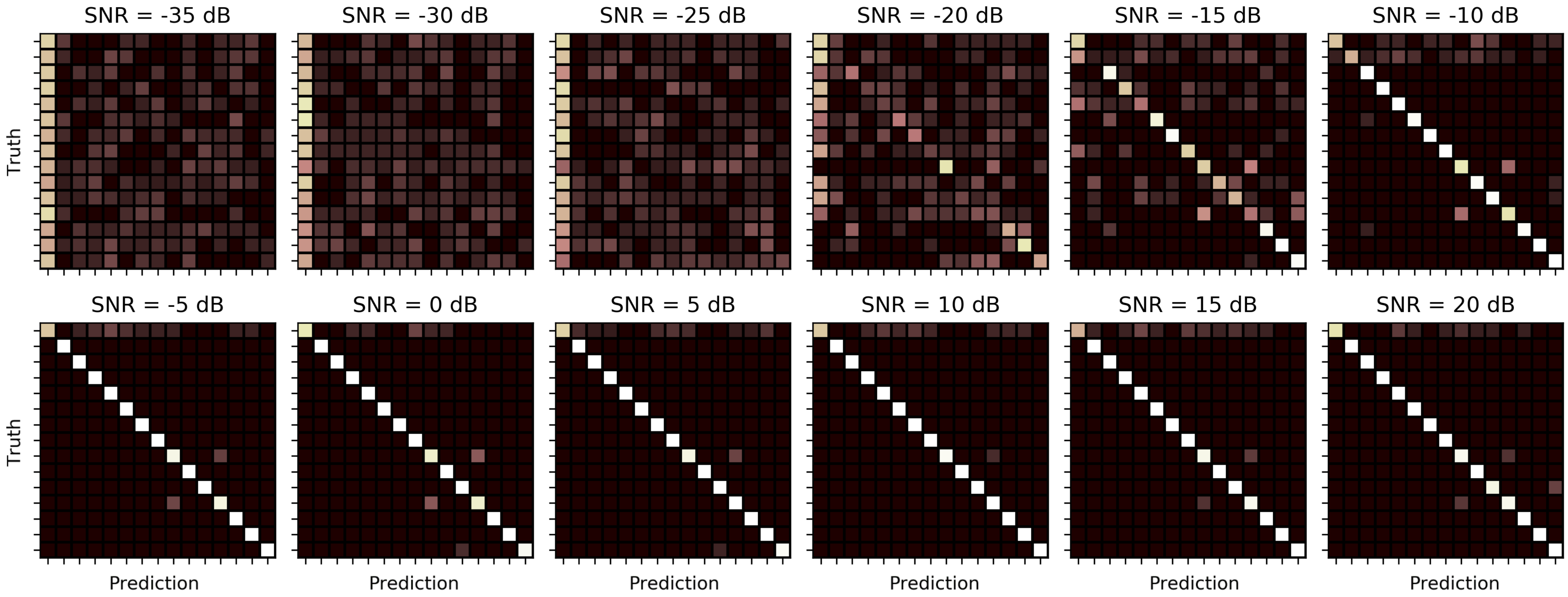
DNN appears to be randomly guessing at low SNR which will create unnecessary requirements on downstream processing

Methodology for Testing False Positive Rate



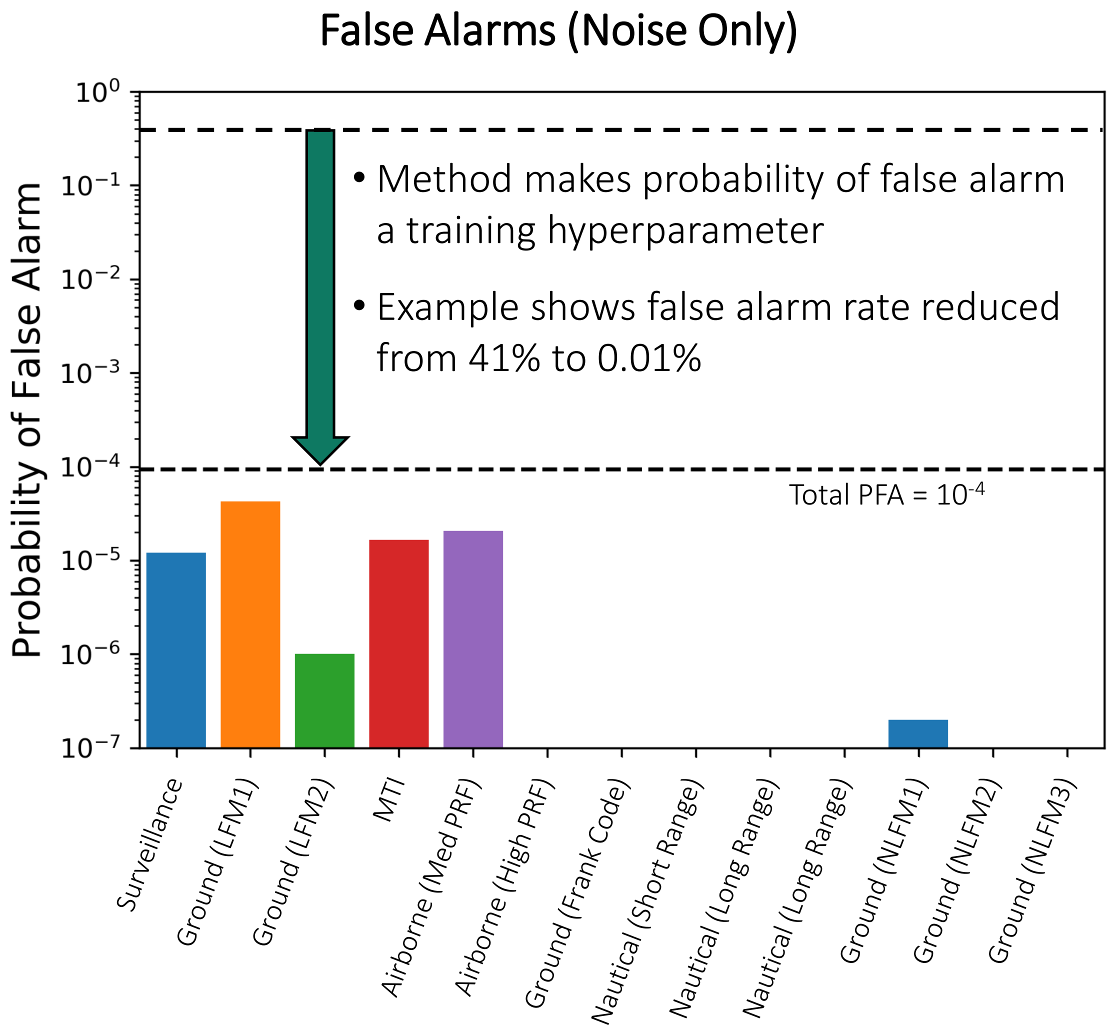


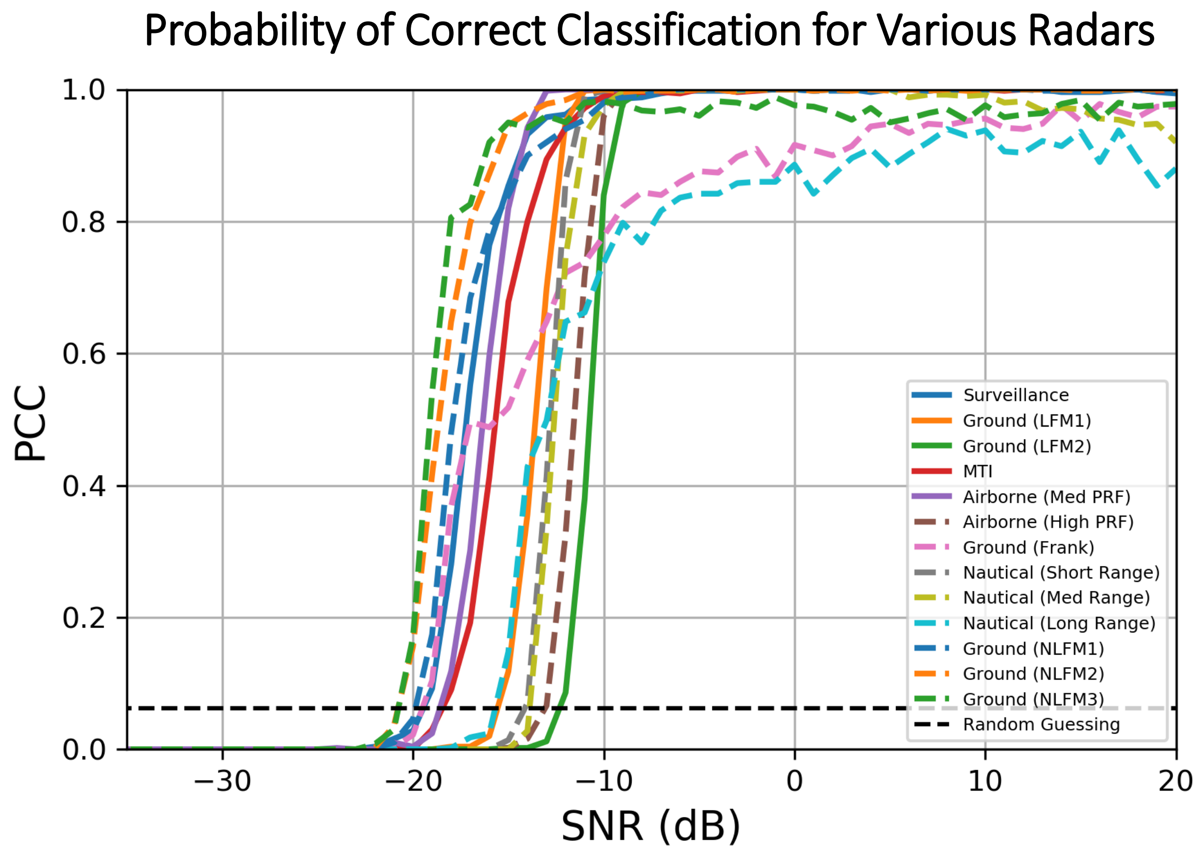
Confusion Matrix and Signal to Noise Ratio



Significant false alarm rate limits algorithm's applicability and creates non-zero probability of correct classification (PCC) at low SNR values

Deepwave Training Method to Reduce False Alarms

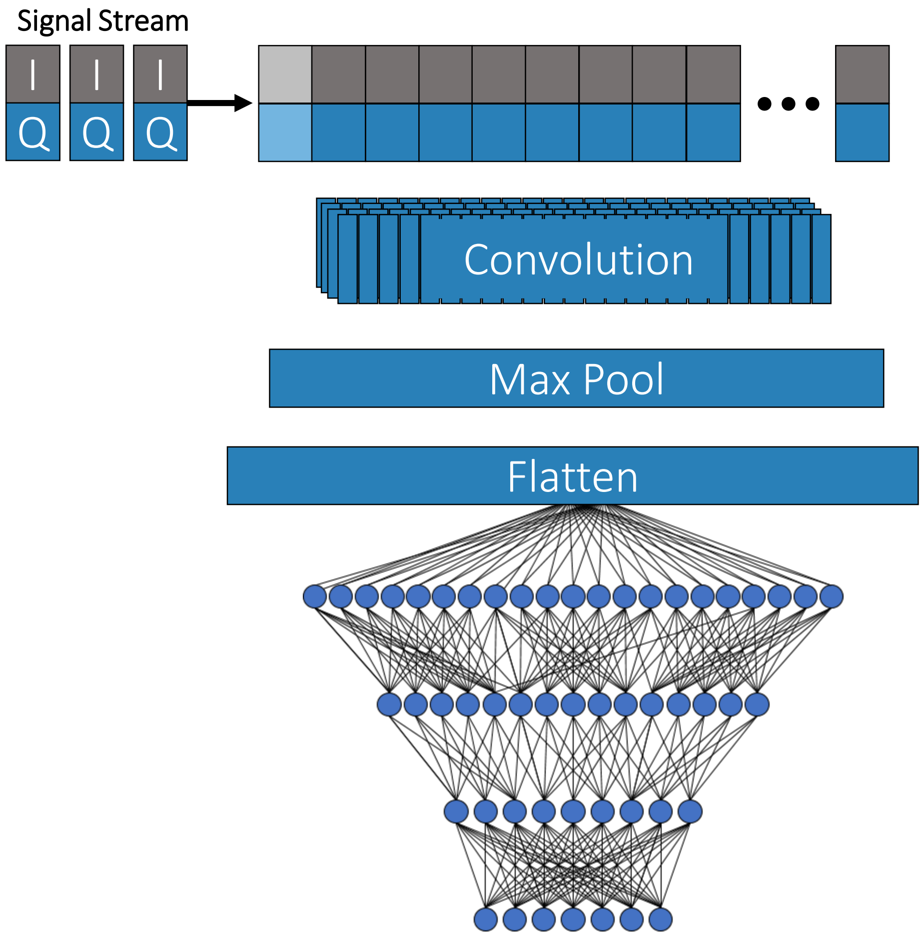




Outline

- Introduction to Deep Learning in RF
- Deepwave's Technology
- Signal Detection and Classification
- Real-time Benchmarks on Embedded GPUs
- Summary

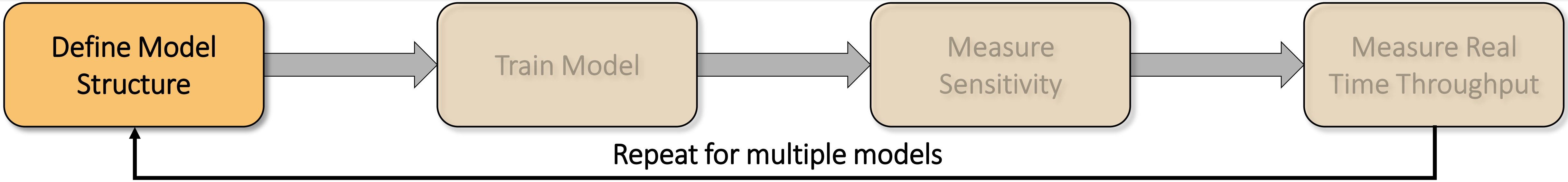
Critical Performance Parameters

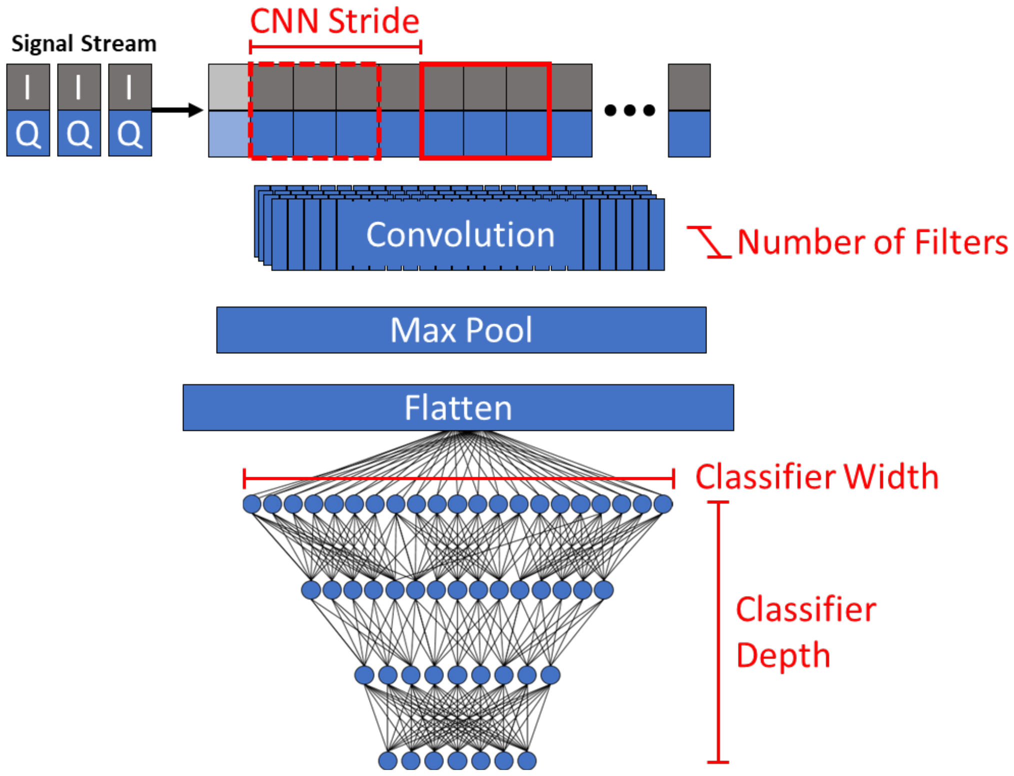


- What makes a DNN model "good?"
  - **High Sensitivity** – detects low powered signals
  - **Low false alarm rate** – minimize false positives
  - **High real time bandwidth**
  - **Low computational requirements**
  - **Low latency**

- Most of these critical performance parameters are adversarial

Performance Benchmarking Test Setup

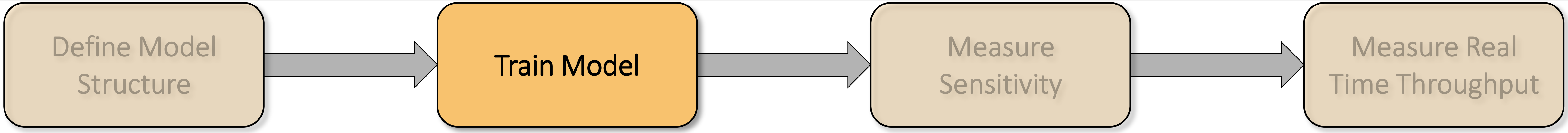




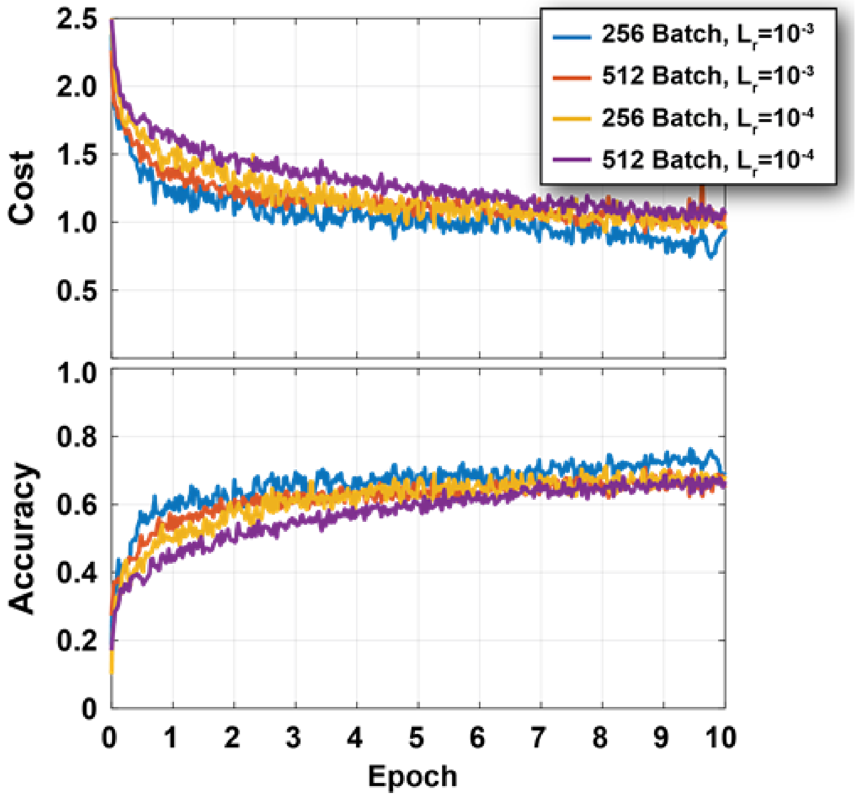
Model Tuning Variables

	Min Val	Max Val	Total
CNN Stride	1	16	9
Number of Filters	4	256	7
Classifier Layer 1 Width	64	128	3
Classifier Layer 2 Width	32	64	3
Classifier Layer 3 Width	0	64	2
Batch Size	1	256	8
Total Model Combinations Tested			728

Performance Benchmarking Test Setup



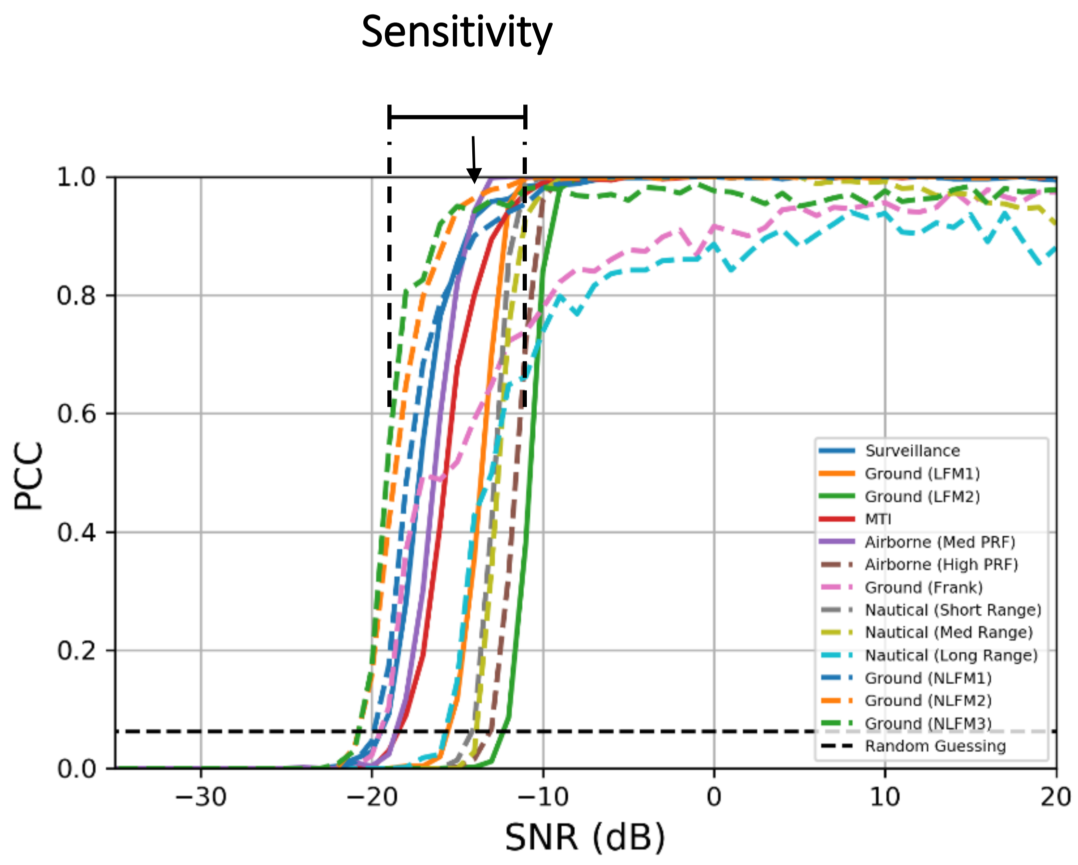
- 1000 training segments per SNR
  - 55 different SNR values
- Softmax cross entropy
- Adam Optimizer
- Quadro GP100 GPU
- Create UFF File for each model



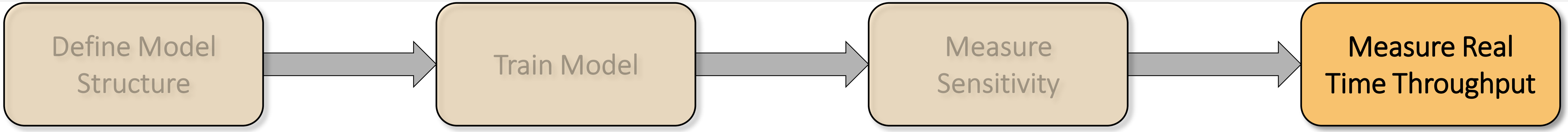
Performance Benchmarking Test Setup

Define Model Structure → Train Model → **Measure Sensitivity** → Measure Real Time Throughput

- Compute receiver operating characteristic (ROC) curve for each model
- Define sensitivity to be where median PCC = 50% for all signal types



Performance Benchmarking Test Setup

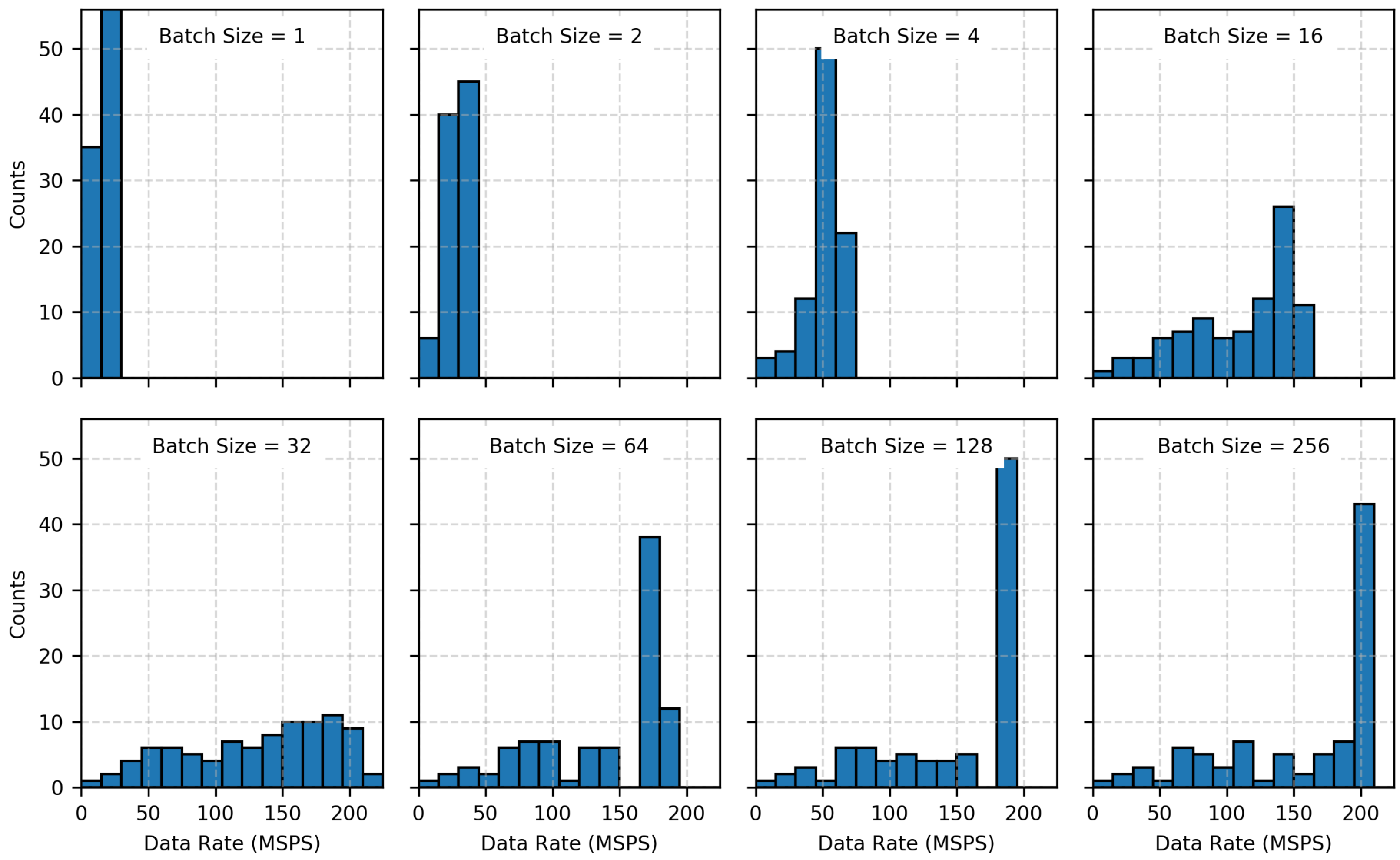


- Create TensorRT PLAN file for each platform tested
- Load signal data into RAM
- Stream unthrottled data to gr-wavelearner

- Measure data rate at two locations:

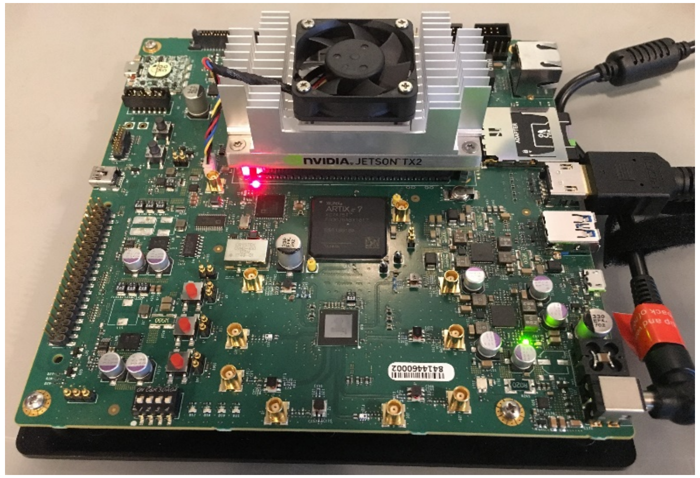
  1. Aggregate data rate for entire process
     - Number of bytes processed / wall time
  2. Computation data rate in work() function
     - Number of bytes process / computation time

Data Rate Benchmark for AIR-T (Tegra TX2)

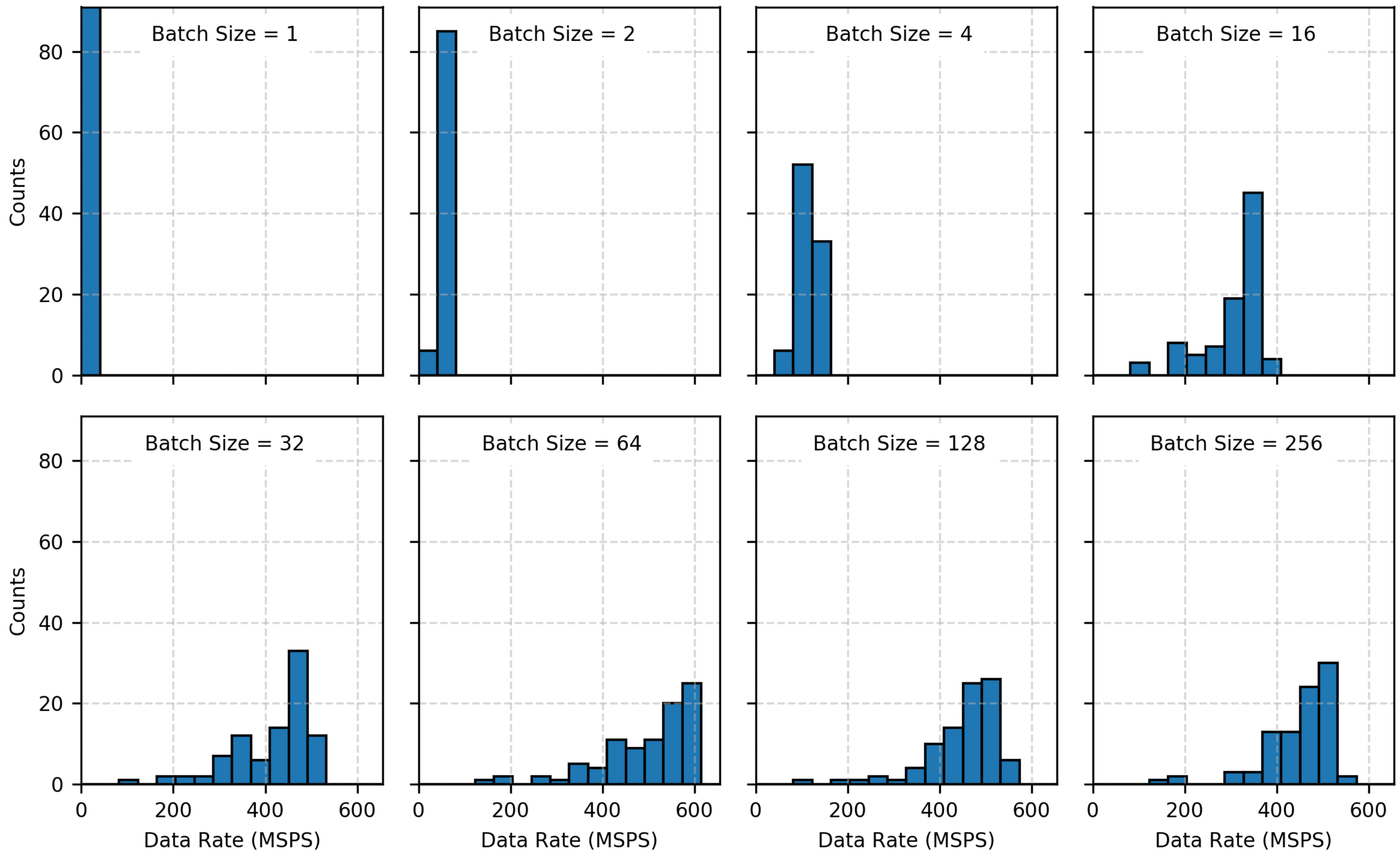


- Tested 91 different CNN classifier models
- Maximum real-time inference data rate for 8 different batch sizes
- Able to achieve 200 MSPS (real samples) with AIR-T

AIR-T



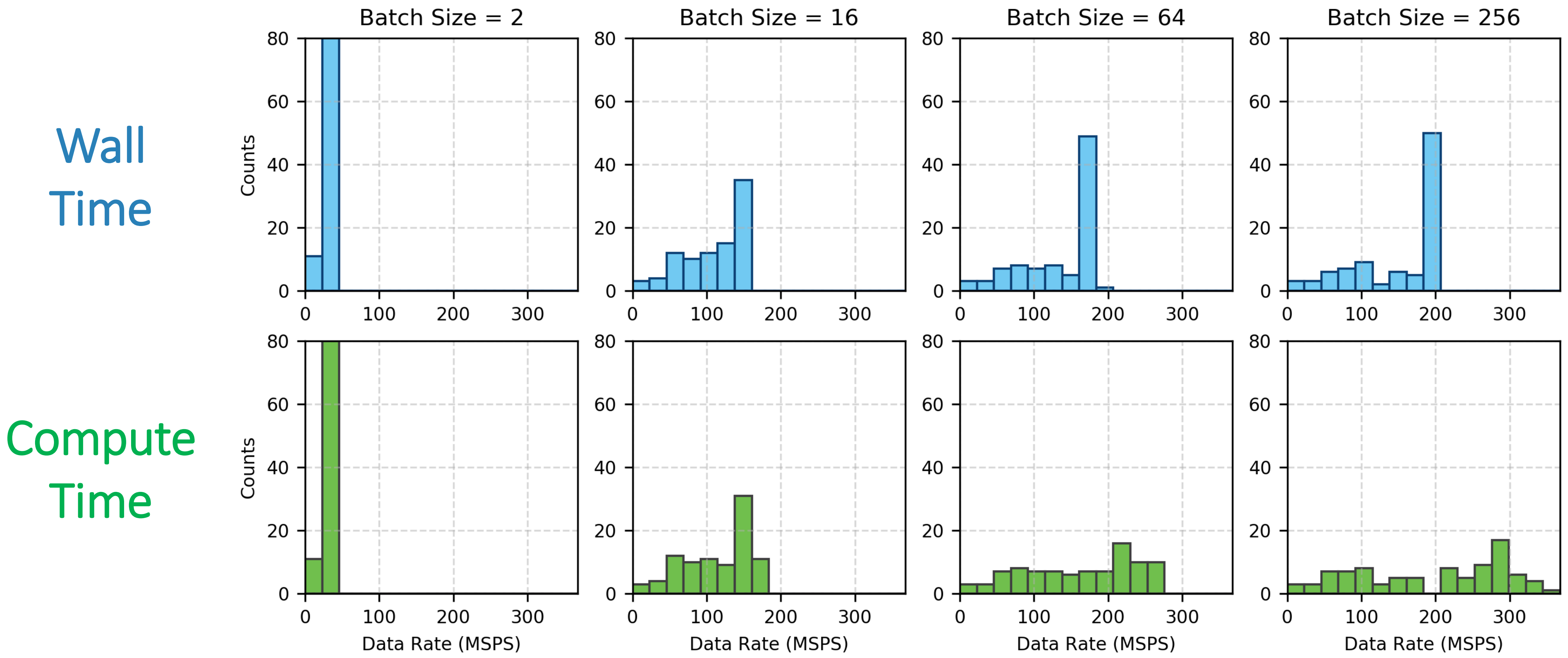
Data Rate Benchmark for Desktop (Quadro P100)



- Tested 91 different CNN classifier models
- Maximum real-time inference data rate for 8 different batch sizes
- Using unified memory will increase throughput

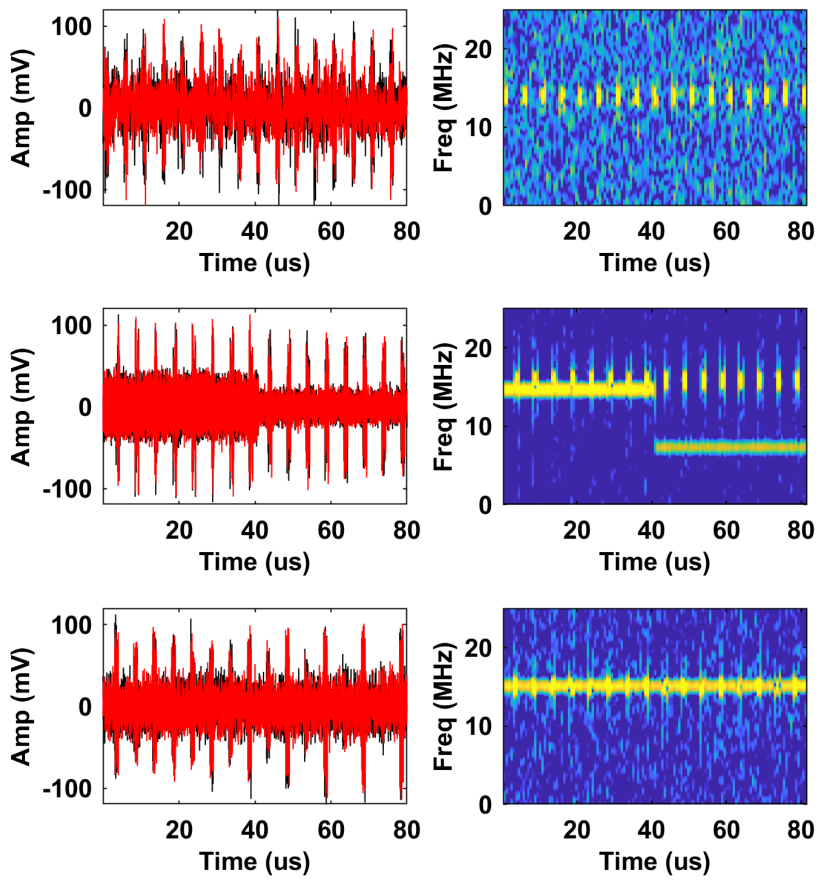
Desktop (GP100)

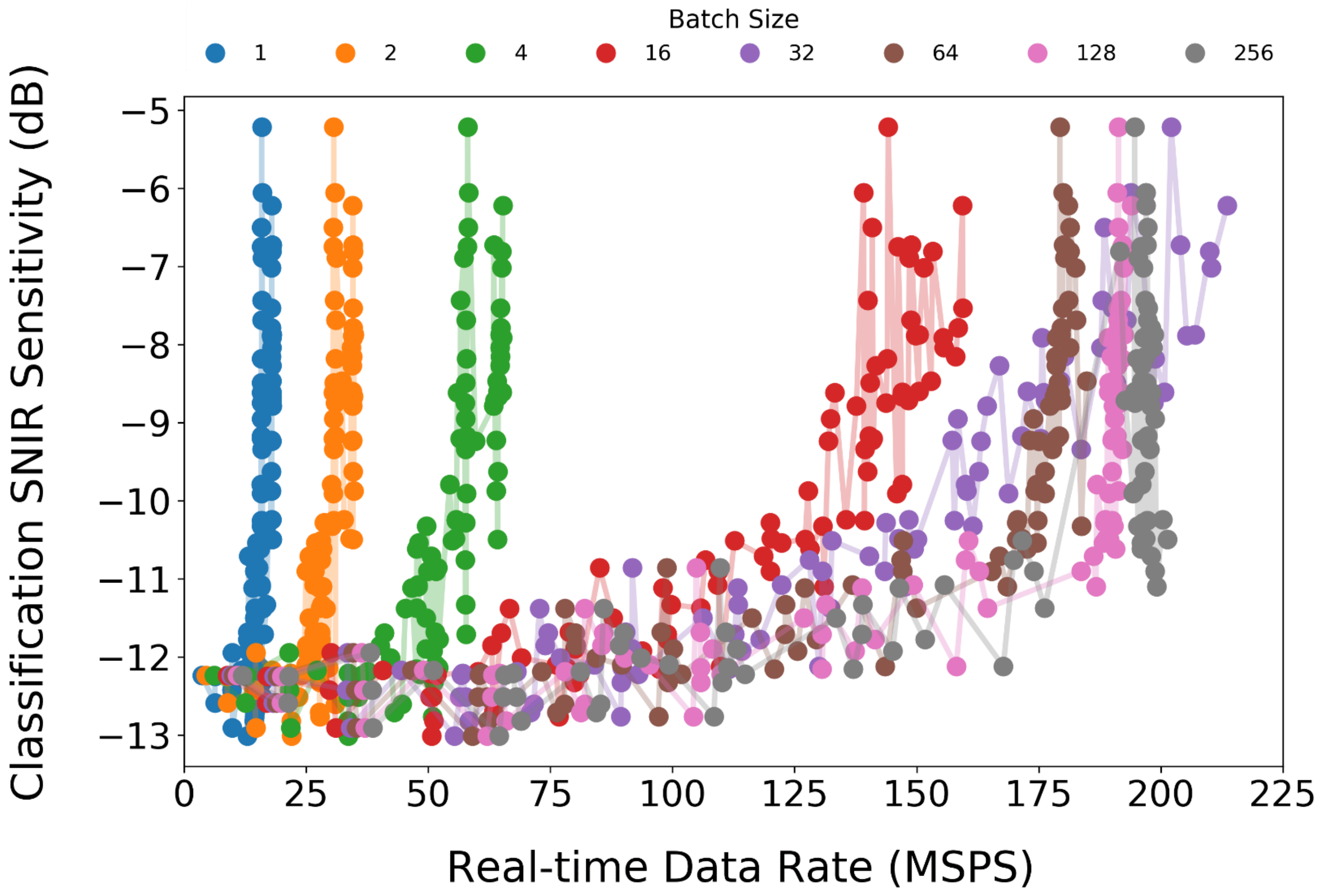
Wall Time vs. Compute Time for AIR-T



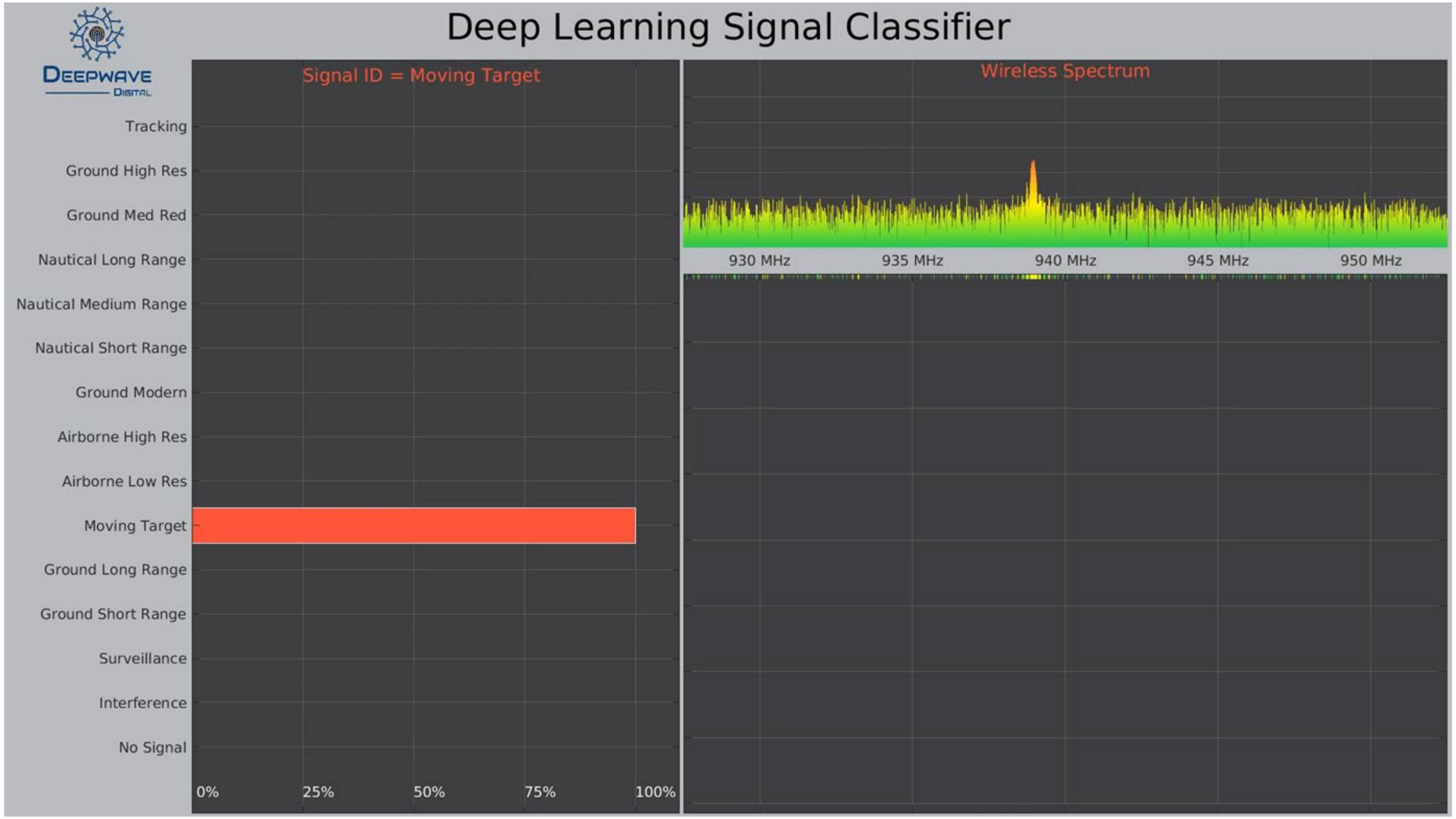
Real time data rate limited by GNU Radio overhead

Model Accuracy Benchmarks





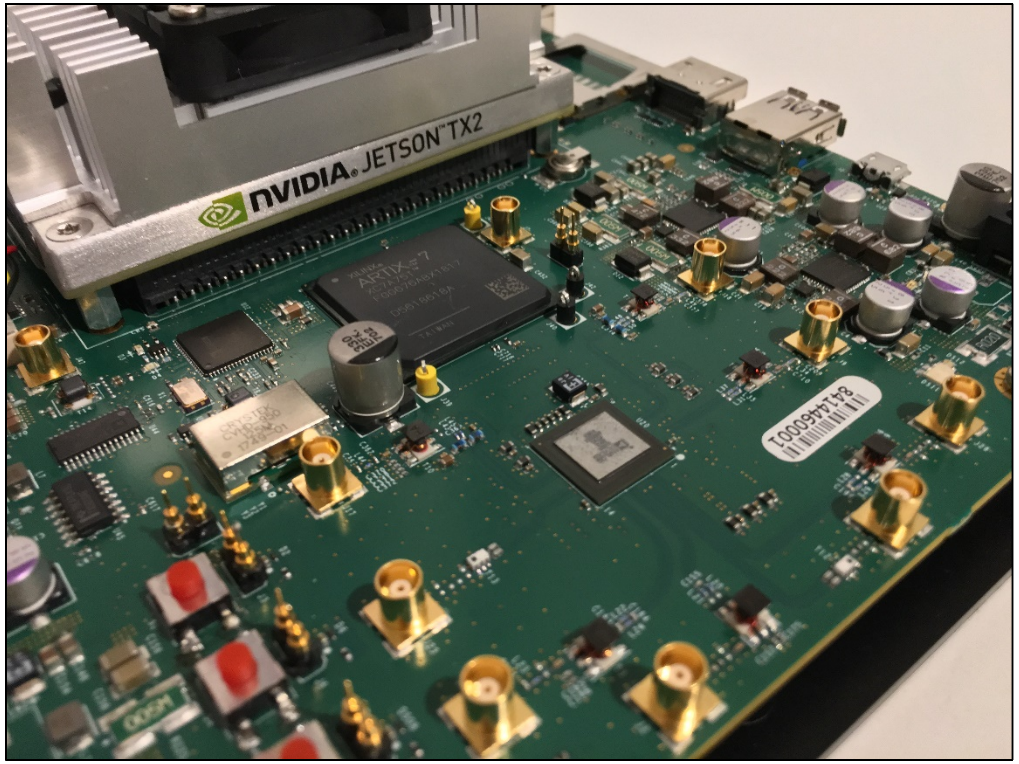
Deepwave Inference Display



Summary

- Deep learning within signal processing is emerging
  - Algorithms may be applied to signal's data content or signal itself
- High bandwidth requirements driving edge solutions
- Deepwave developed AIR-T
  - Edge-compute inference engine with MIMO transceiver
  - FPGA, CPU, GPU
- GR-Wavelearner
  - Open source inference engine for signal processing
  - Available now on our GitHub page

More info at www.deepwavedigital.com/sdr



- Benchmarking analysis demonstrates AIR-T with GR-Wavelearner capable of signal classification inference at 200 MSPS real-time data rates
  - Improvements likely in future release

DEEPWAVE

DIGITAL

info@deepwavedigital.com