This research was developed with funding from the Defense Advanced Research Projects Agency (DARPA). The views, opinions, and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

CASPER: Security Monitoring Using Unintended RF Emissions

GNU Radio Conference 2018

Joshua Morman, Rob Miller, Joe Liberti
Scott Alexander, Simon Tsang, Marc Pucci
Christine Hung, Chris Mesterharm, Euthimios Panagos, Isil Sebuktekin

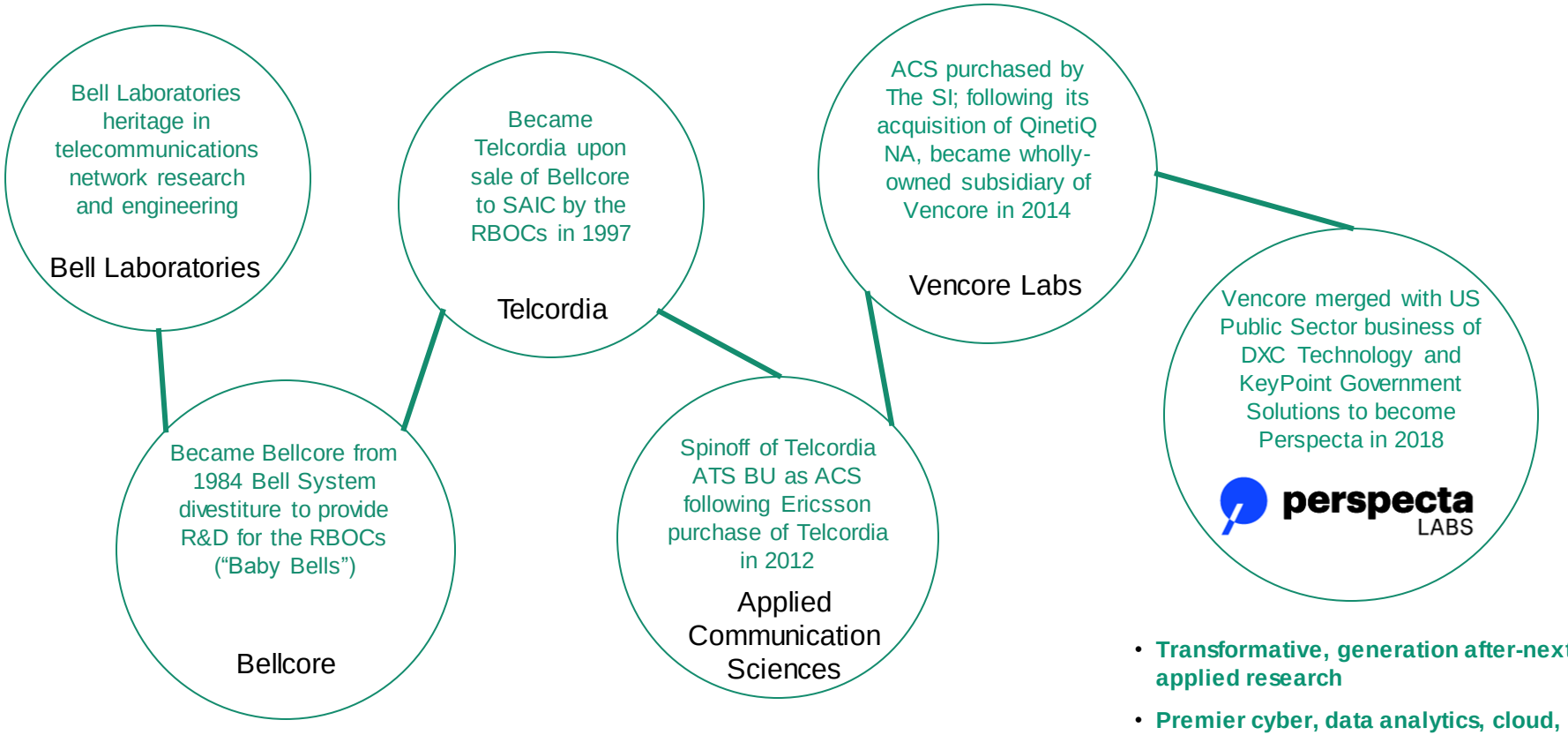
perspecta LABS

Overview

- Unintended Processor Emissions
- CASPER System Overview
- Collection Methods
- Machine Learning
- Real-time streaming system
  - Demo
- Array Processing
- Conclusion and Future Direction

Perspecta Labs

Who Are We?



- **Transformative, generation after-next applied research**
- **Premier cyber, data analytics, cloud, quantum, mobility, and advanced networking capabilities**

Background

Motivation

- IoT and industrial device vulnerabilities
  - Many IoT devices severely lacking in security controls
  - Susceptible to a variety of threats
  - Traditional approach focuses on digital/cyber monitoring solutions
- If a device becomes compromised, or infected with malware, how do we detect this without access to the processor itself?

Source: https://www.businessinsider.com/what-hackers-can-do-to-our-power-grid-2014-11

businessinsider.com

What Hackers Can Do To Our Power Grid

Jonathan Pollet, Contributor

10-13 minutes

Source: https://pennyelectric.com/blog/smart-meter-controversy-las-vegas-need-know/

cnbc.com

Suddenly hot smart home devices are ripe for hacking, experts warn

Jennifer Schlesinger, Andrea Day

10-12 minutes

Will 2017 be the year your home becomes under attack from cyber criminals?

Source: https://www.cnbc.com/2016/12/25/suddenly-hot-smart-home-devices-are-ripe-for-hacking-experts-warn.html

Background

Side Channel Attacks

- Well known phenomenon and subject of much security research
- Information gained by leveraging unintentional aspects of a system
  - Electromagnetic leakage
  - Power consumption
  - Sound
  - Timing
  - Differential Fault Analysis
- This information can be exploited to reveal information, such as RSA keys

Source: https://www.tau.ac.il/~tromer/acoustic/

RSA Key Extraction via Low-Bandwidth Acoustic Cryptanalysis

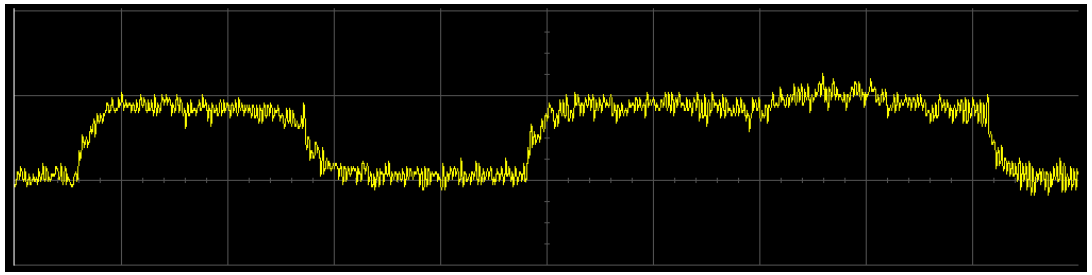
Daniel Genkin — Technion and Tel Aviv University

Adi Shamir — Weizmann Institute of Science

Eran Tromer — Tel Aviv University

assisted by Lev Pachmanov and numerous others

Summary



Source: https://en.wikipedia.org/wiki/Side-channel_attack#/media/File:Power_attack.png

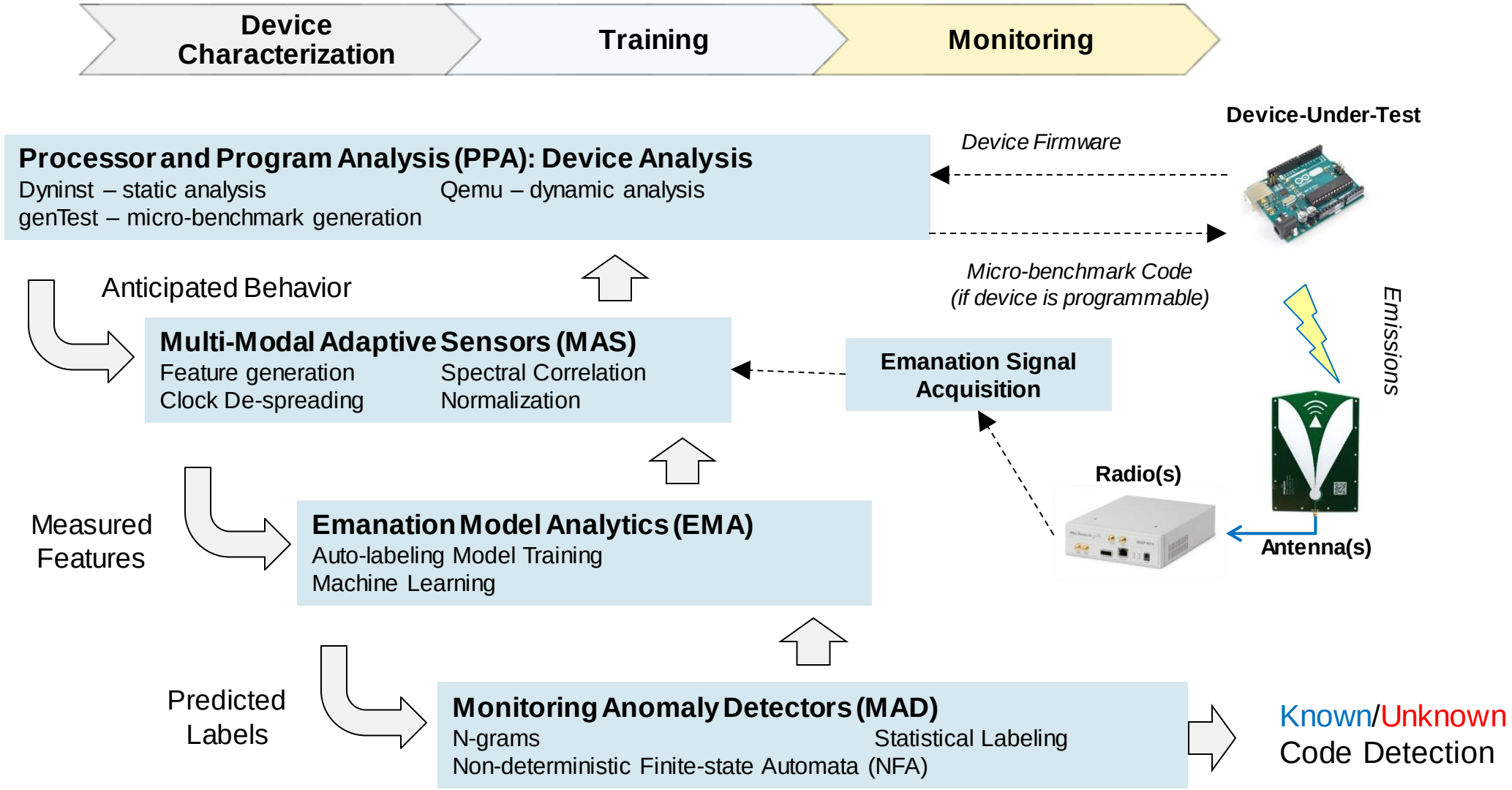
DARPA LADS

Leveraging the Analog Domain for Security

- **Goal**: **Develop new cyber techniques in digital devices by monitoring the unintended analog emissions across different/multiple modalities**:
  - Tracking fidelity vs. device complexity
  - Fidelity: Known/unknown code, control flow tracking, instruction tracking, …
- **Output**: **Monitoring devices; network architectures; algorithms for mapping digital artifacts to analog emissions**
- **Methodology**:
  - Identify and quantify useful analog signals: develop predictive models
  - Map device firmware, configuration, and data to cyber-relevant analog emissions model: unknown firmware & configuration
  - Boost signal via software and/or analog component modifications
  - Reconcile tracked device emissions with emissions model
  - Cooperative sensing and tracking

Source: DARPA LADS Proposers Day Slides 1 Oct 2015

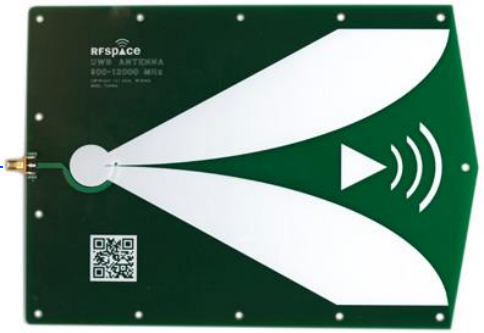
CASPER System Overview



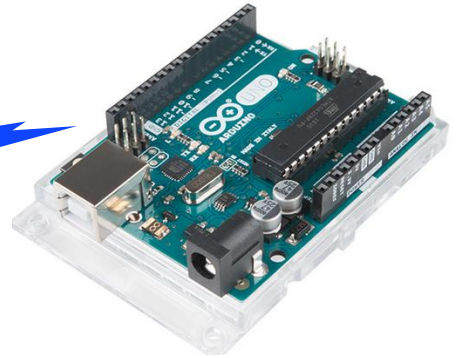
Collection Methods

Source: ettus.com

Source: rfspace.com

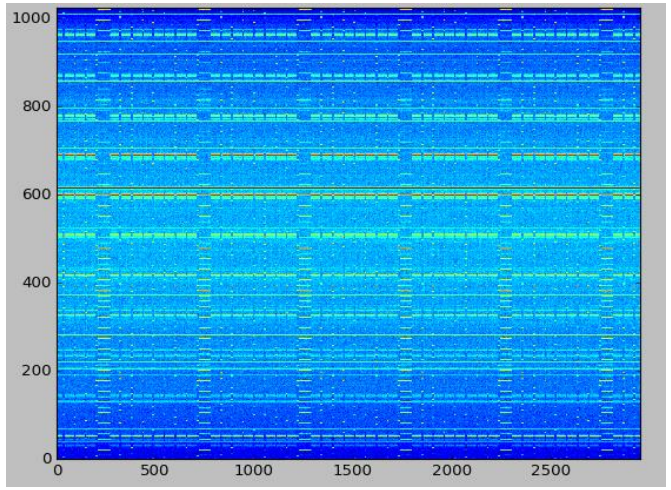


Source: sparkfun.com

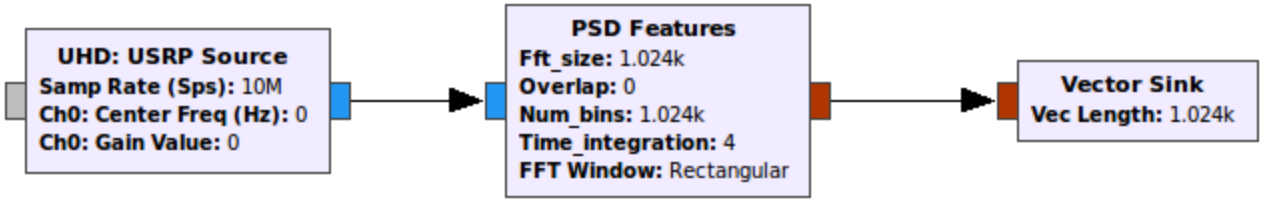


- RF Modality
  - Use USRP and gnuradio to capture emanations
  - Antenna
    - Vivaldi Slot Antenna
  - Magnetic Near Field Probe
    - Placed right next to processor, captures magnetic field fluctuations

- Power Modality
  - Tap directly off Vcc pin
  - Toroidal Coil on power line
  - USRP with LFRX



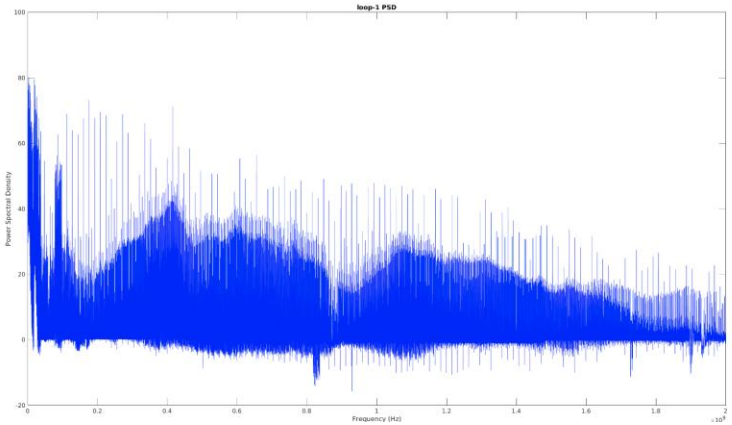
- ***Quick demonstration of how changing programs on processor impacts RF emanations***



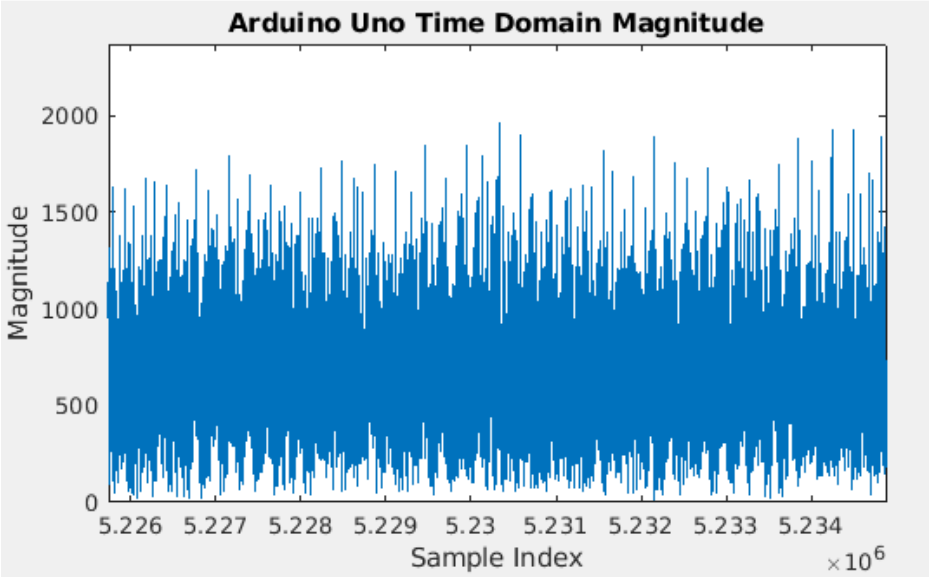
Processor Emission Model

What is causing the emanations?

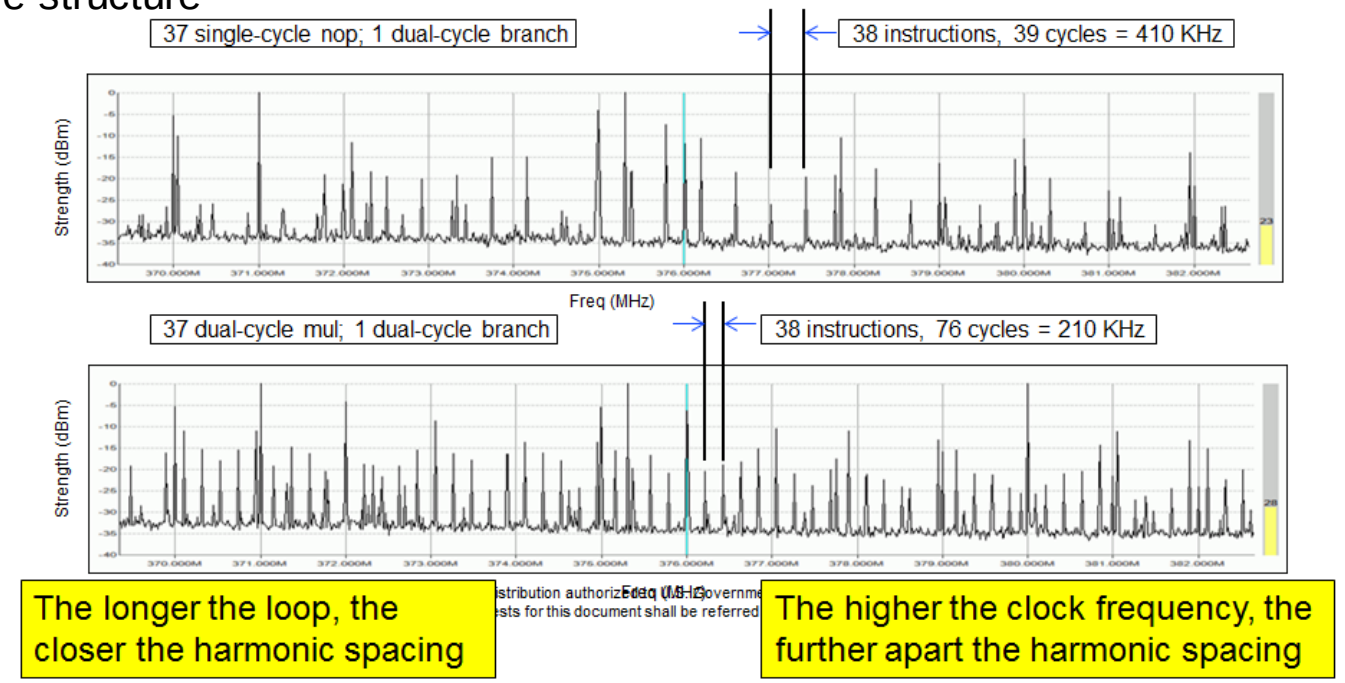
0-2GHz –peaks show up across the spectrum



- Very noticeable emanations show up at harmonics of the clock frequency
  - Program loops are a significant source of CPU device emissions
  - Repetitive state transitions at the loop frequency create harmonics throughout the spectrum
  - Spike separation = clock cycles per second / loop length in cycles
  - Spikes vary in magnitude – emissions have structure



16 MHz Arduino Uno emissions measured at 376 MHz



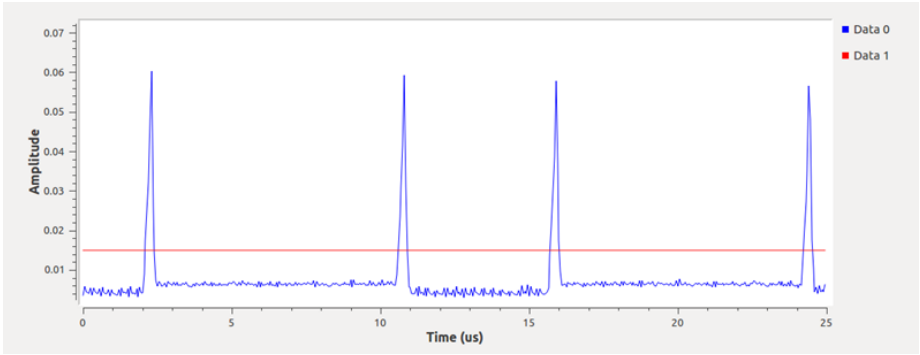
The longer the loop, the closer the harmonic spacing

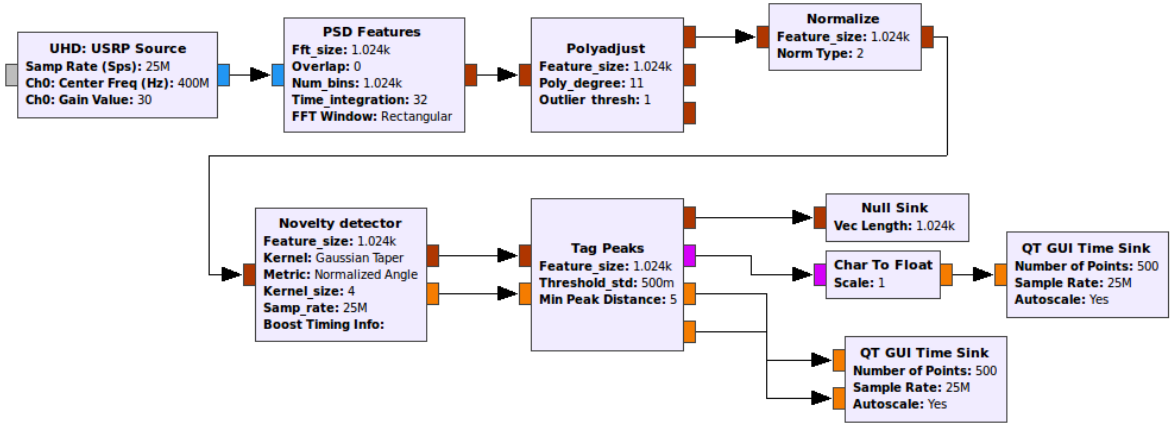
The higher the clock frequency, the further apart the harmonic spacing

Signal Features

Feature Extraction

- For most of the emanations we have observed, time averaged FFTs estimating the Power Spectral Density capture the necessary information to classify the program modes
  - Very dependent on the device/program being evaluated
- We have evaluated other signal features with varying results on various devices
  - Wavelet/Wavelet Leader transforms
  - Cepstral processing
  - Time domain statistics (Zero crossing, amplitude histograms, etc.)
  - Autoencoders
  - Burst detection
  - Spectral correlation
  - Data dependent basis functions
  - PCA/LDA reduction
- Normalization/Equalization/Spectral Baseline Removal
  - Mitigate measurement/device differences
- Novelty/Change Detection

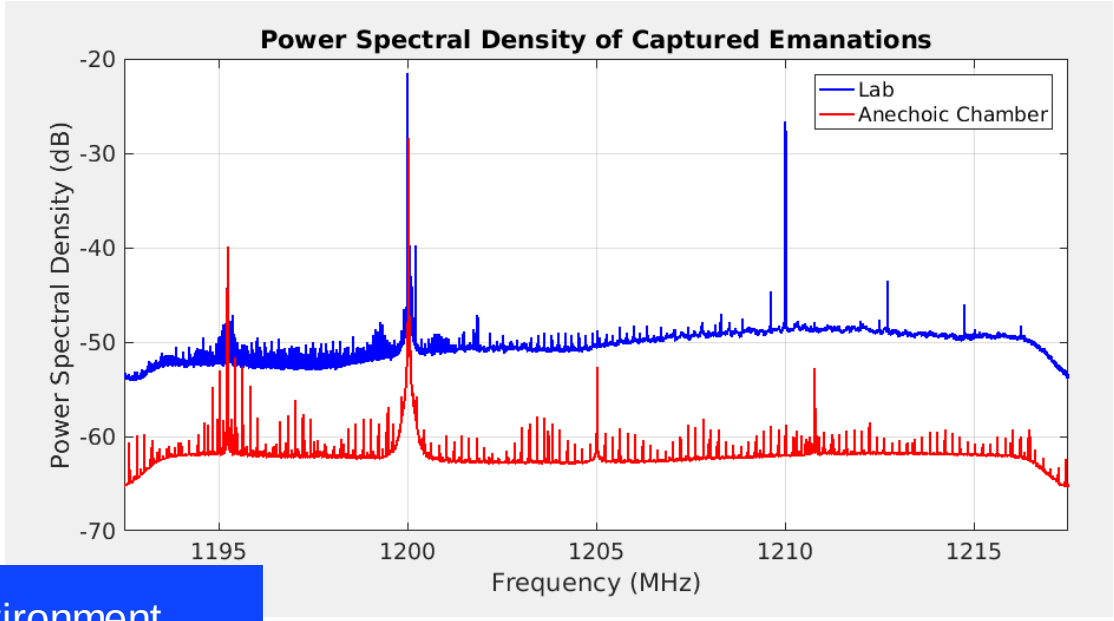




Automated Band Scan

Finding the useful frequencies

- At a standoff distance, low power emissions can be overwhelmed by environmental noise – e.g. radio/TV stations, cell towers, EMI, etc.
  - We will introduce other methods to deal with this later
- Devices emanate across the spectrum to a varying degree
  - Because of the clock harmonics
- Find the “best” spot to tune the radio
- What is “ground truth”?
  - How do we know what is the best place to tune
- What is our measure of information

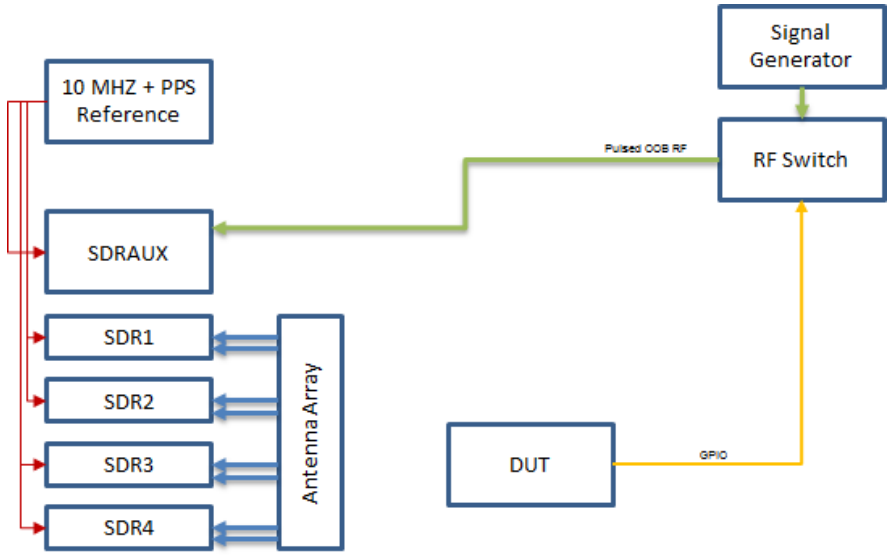


In a typical lab environment, emanations that scream in an anechoic chamber are overwhelmed by EMI

Automated Band Scan

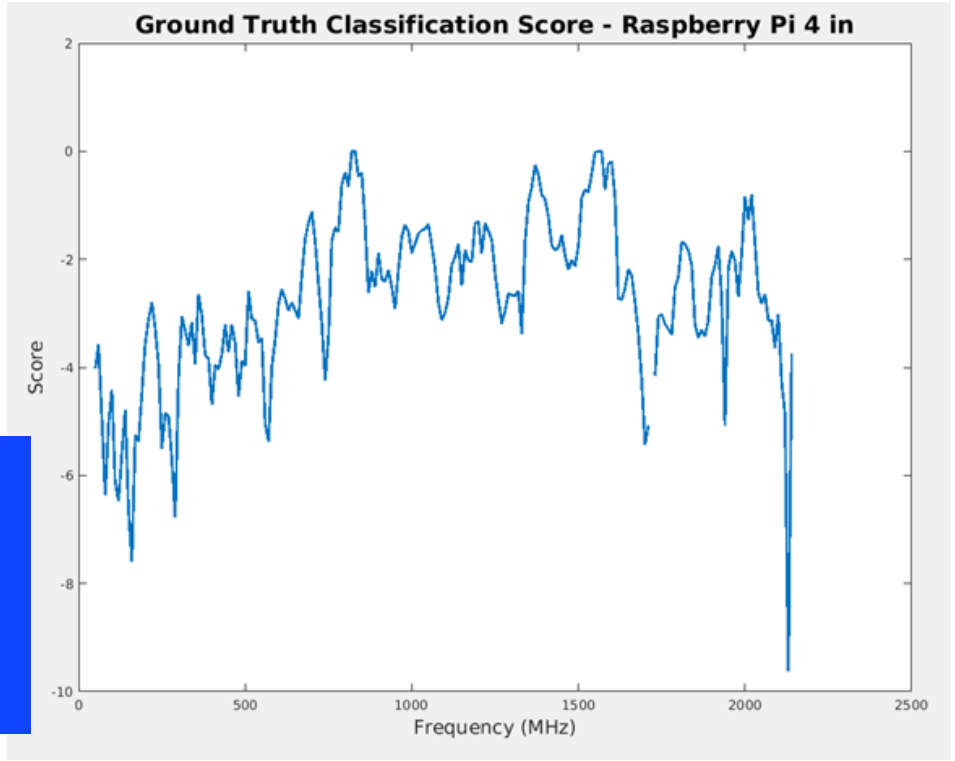
Establishing Ground Truth

- With an instrumented setup, we can parse program sections with processor GPIO and capture as an out of band RF signal



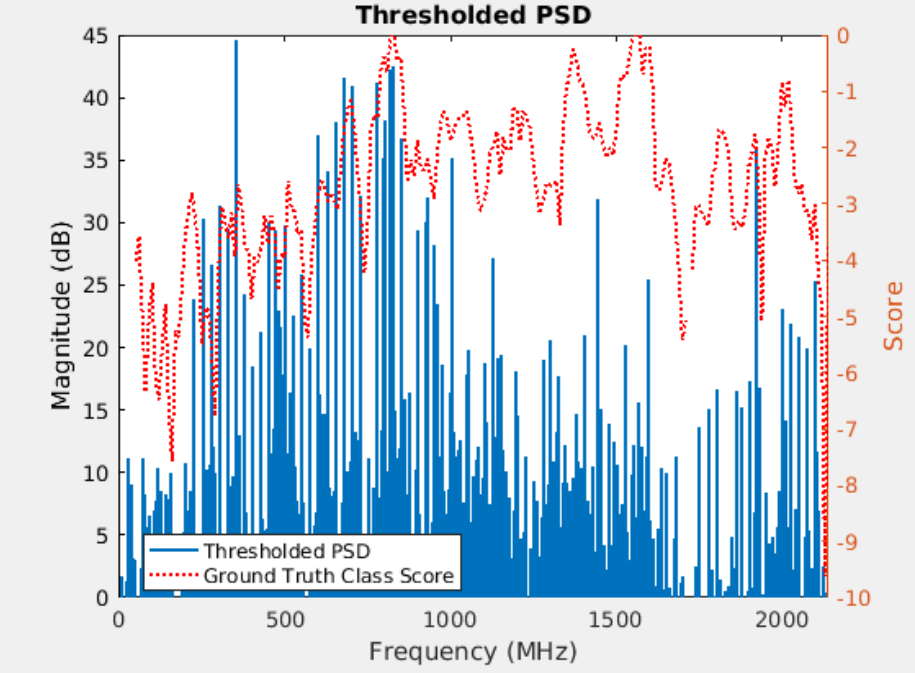
- Establish a ground truth classification score which is related to the smallest time averaging where a classification threshold is met

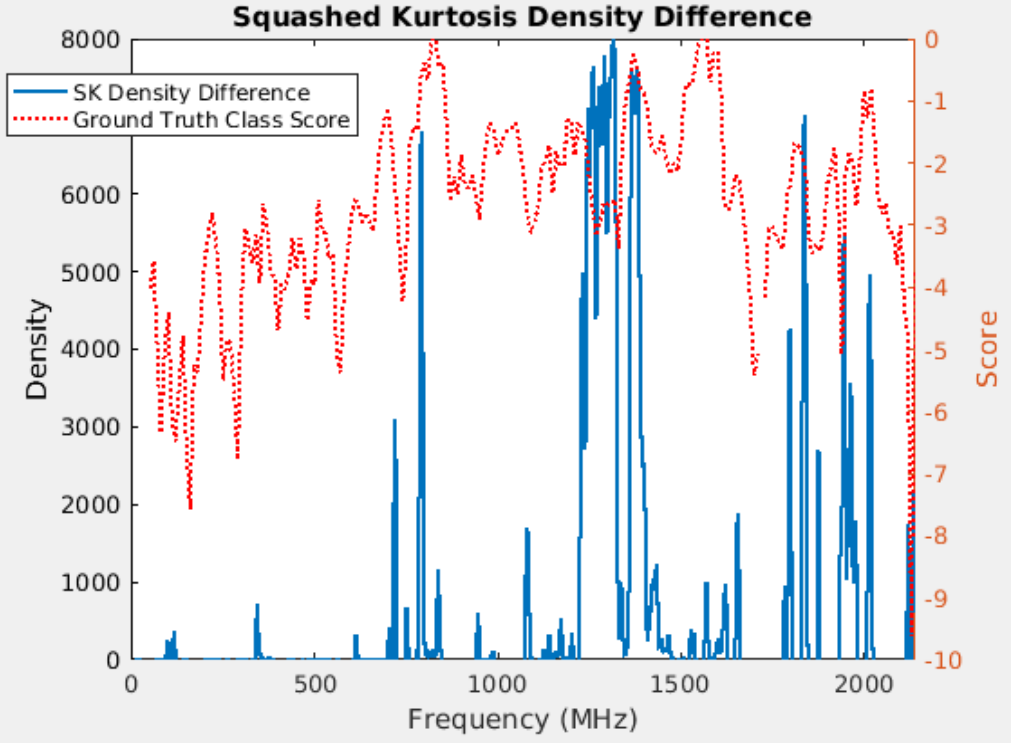
Baseline program on Raspberry Pi – 4 states - FFT, buffer fill, String Search, SHA



Automated Band Scan

- Capture signal statistics over a time window for a 25 MHz band stepping across 2+ GHz
  - Capture with device off first to establish a detection threshold
- PSD alone does not peak at maximum ground truth score
- Spectral kurtosis peaks up at frequencies with non-Gaussian properties





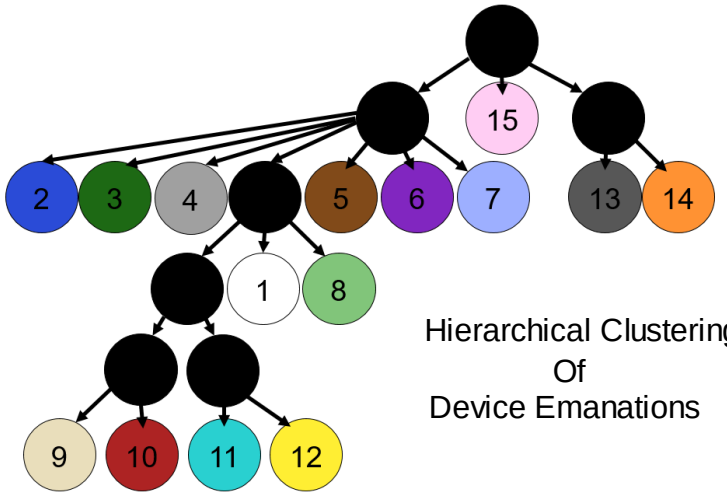
Machine Learning

Labeling Unlabeled Data

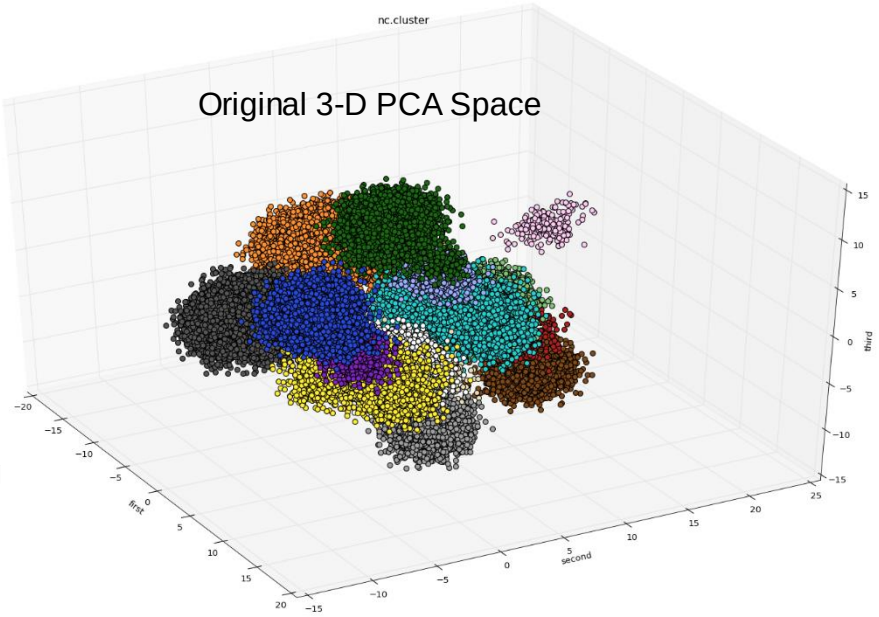
Goal: Automatically label the features produced by the signal processing component

- Labeled data is needed for training CASPER and to develop device models
- Use PCA and DBSCAN to identify clusters of program modes
  - For training some DBSCAN parameter selection is currently used to label clusters though default parameter often work well
  - For many devices this is sufficient, but often the clusters produced are given their non-uniform densities and overlapping nature in N-D space

- Our new technique does a hierarchical clustering that reapplies PCA and DBSCAN on subsets of points
  - This solves the density problem by selectively removing clusters that cause problems for PCA and DBSCAN
  - This allows a richer set of dimensions while allowing 3D visualization for individual hierarchy nodes

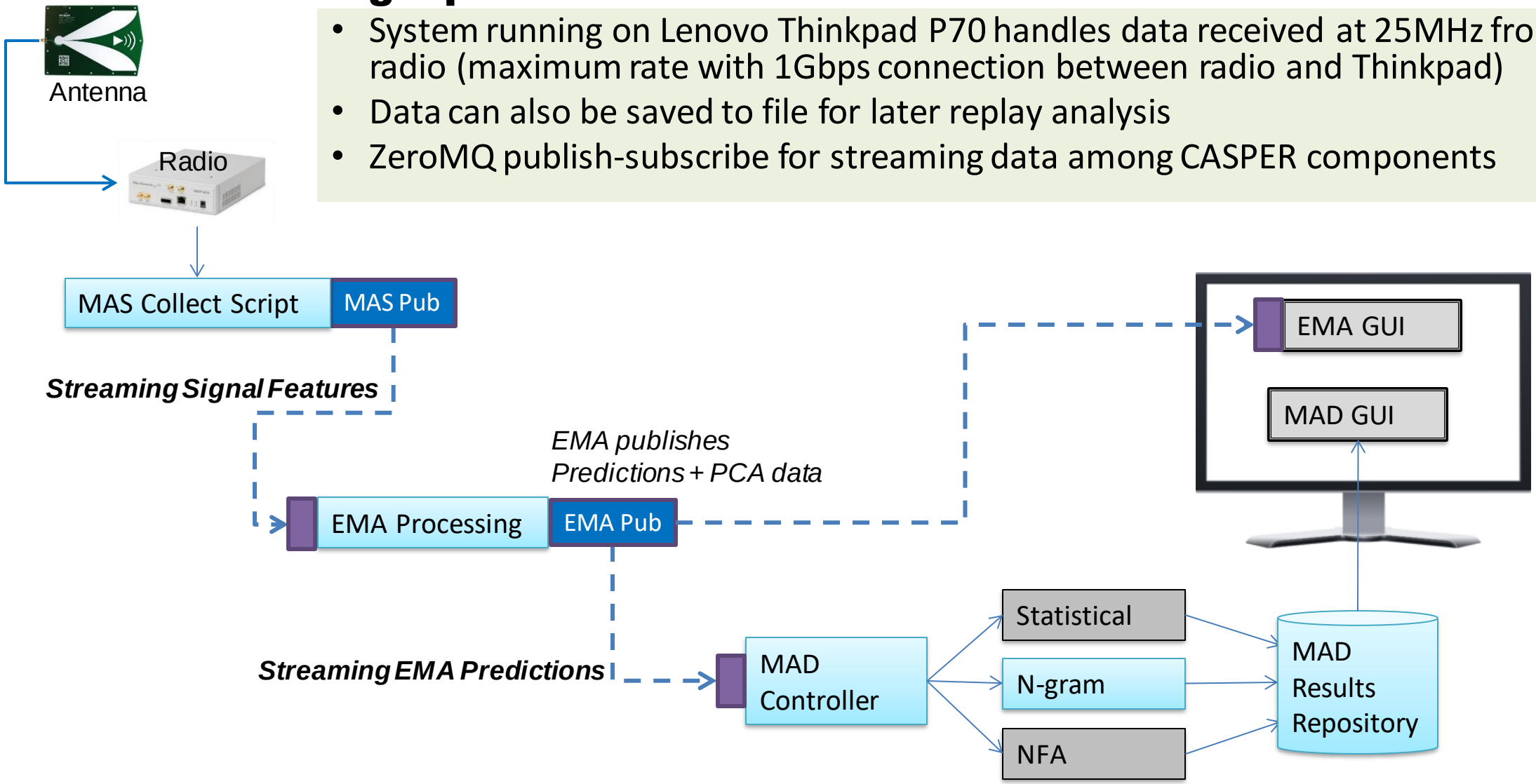


Hierarchical Clustering
Of
Device Emanations



Real-time Streaming Operation

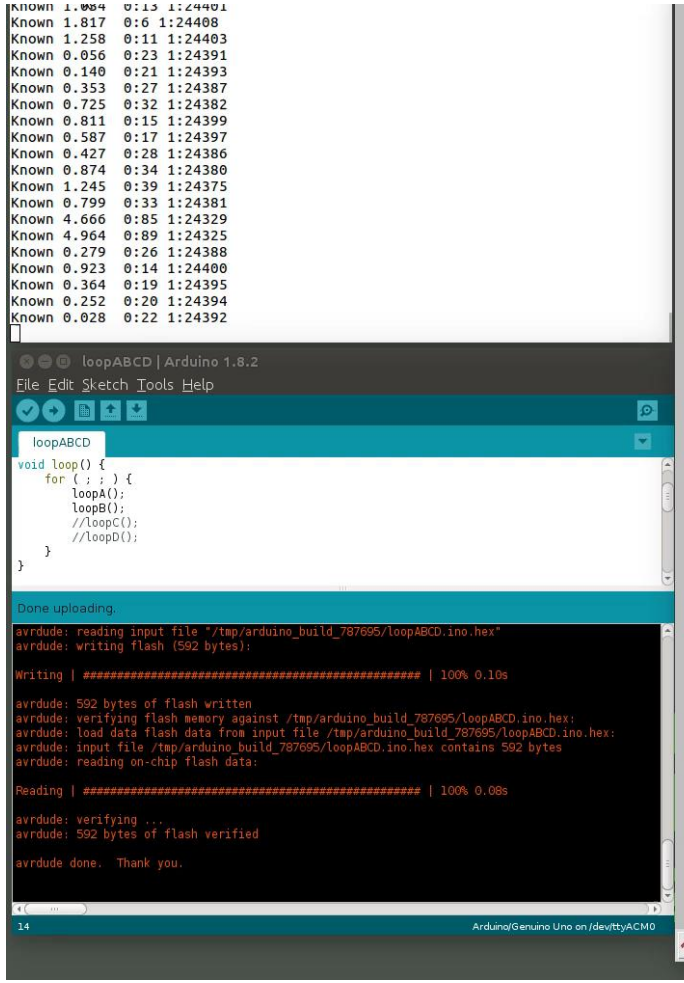
- System running on Lenovo Thinkpad P70 handles data received at 25MHz from radio (maximum rate with 1Gbps connection between radio and Thinkpad)
- Data can also be saved to file for later replay analysis
- ZeroMQ publish-subscribe for streaming data among CASPER components

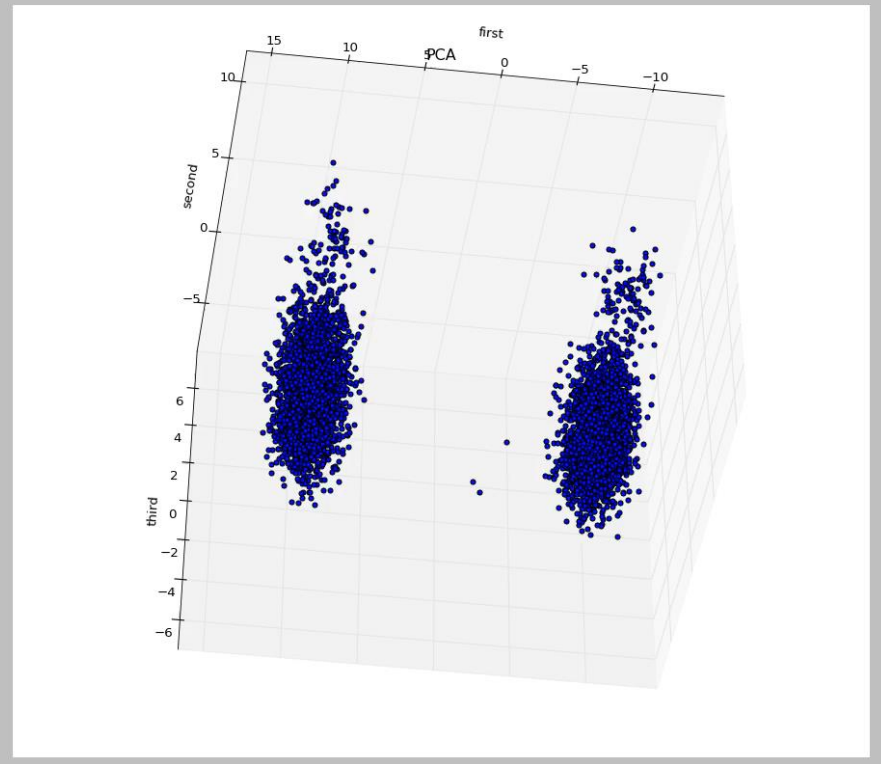


Streaming MAD Unknown Alerts and Detection Statistics

Real-time Streaming Operation

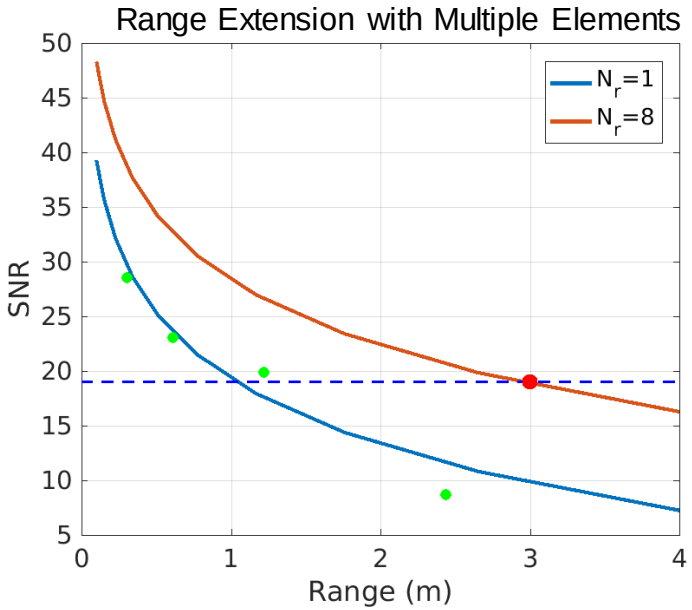
Demo





Array Processing

Introduction



- Why do we need it?
  - As a general technique to suppress background noise and interference and extract a desired signal
  - The power of unintended emissions from processors are very low
- How does it work
  - We apply a weight vector w(n) to each array element that emphasizes some signals and nulls interfering signals
  - The trick is to find the optimal weight vector w(n) $\sum^{N} w_n r_n = \sum^{N} w_n soi_n + w_n noise_n + w_n int_n$
- The net benefits of array processing are:
  1. Extend range over what we could do with a single element
  2. Unlike using a single big directional antenna, when using an array we don't need to know where the signal we want is coming from – we can let the array signal processing pull it out
  3. Mitigate interference (everything from a television station that operates in a chunk of spectrum that's useful to monitor, to interference from other unintentional radiators in the room and the cell phone in your pocket)

Find w(n) that emphasizes soi, minimizes noise and interferer from the received signal r(n)

Array Processing

Challenges

- Challenges with array processing unknown signals
  - Uncalibrated array
    - For calibrated array, these weights can be calculated based on direction of arrival
  - Indoor multipath
  - Unknown direction of signal and interferers
  - ***Emanations don't have training sequences or well defined modulation characteristics***

- We need a blind technique
  - Interference Baselining
  - Independent Component Analysis
  - ***Adaptive Event Processing***
- There is a defining characteristic we *can* take advantage of:
  - Switching on and off at certain frequency subcarriers

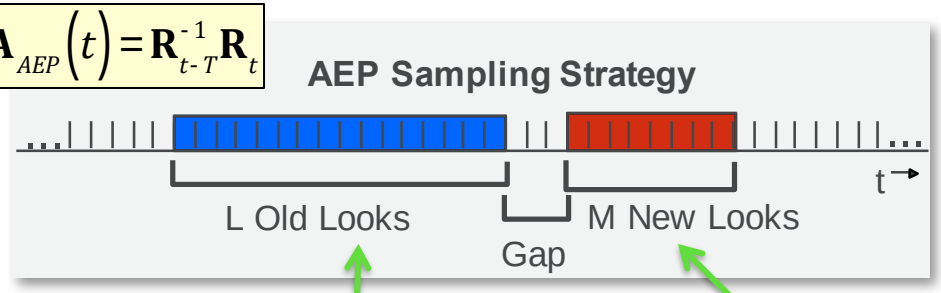
Array Processing

AEP Theory

We use adaptive event processing (AEP) to detect when signals change over time by looking at the time-varying spatial covariance for each subcarrier.

For the case of $\boldsymbol{g}^h\boldsymbol{h}=0$, we can form the Adaptive Event Processing (AEP) metric for each subcarrier:

$$\mathbf{A}_{AEP}(t) = \mathbf{R}_{t-T}^{-1}\mathbf{R}_t$$

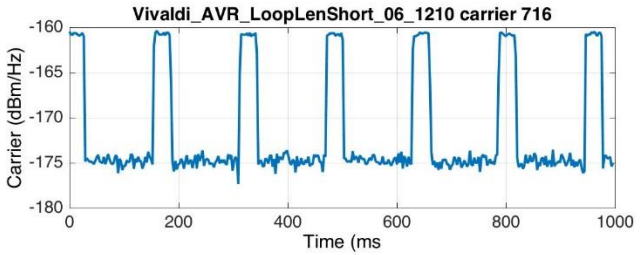


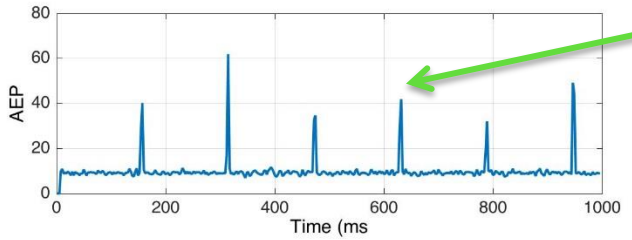
$$\mathbf{r}_1 = \mathbf{g}s_g + \mathbf{n} \qquad \mathbf{r}_2 = \mathbf{h}s_h + \mathbf{g}s_g + \mathbf{n}$$

$$\mathbf{R}_2^{-1}\mathbf{R}_1 = \begin{bmatrix} \mathbf{u}_1 & \mathbf{U}_{N1} \end{bmatrix} \begin{bmatrix} \lambda_{10}^{-1} & 0 \\ 0 & \Lambda_{N1}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{u}_1^H \\ \mathbf{U}_{N1}^H \end{bmatrix} \left(P_h\mathbf{h}\mathbf{h}^H + P_g\mathbf{g}\mathbf{g}^H + \sigma_n^2\right)$$

$$= \mathbf{I} + \frac{P_h}{\sigma^2}\mathbf{U}_{N1}\mathbf{U}_{N1}^H\mathbf{h}\mathbf{h}^H$$

The principal component of $\mathbf{A}_{AEP}(t)$ gives an estimate for $\mathbf{h}$, in the null space of $\mathbf{g}$. So now we only need to look for the peaks in the AEP metric, indicating that a change has occurred. At that point in time, the AEP matrix has the structure shown above.





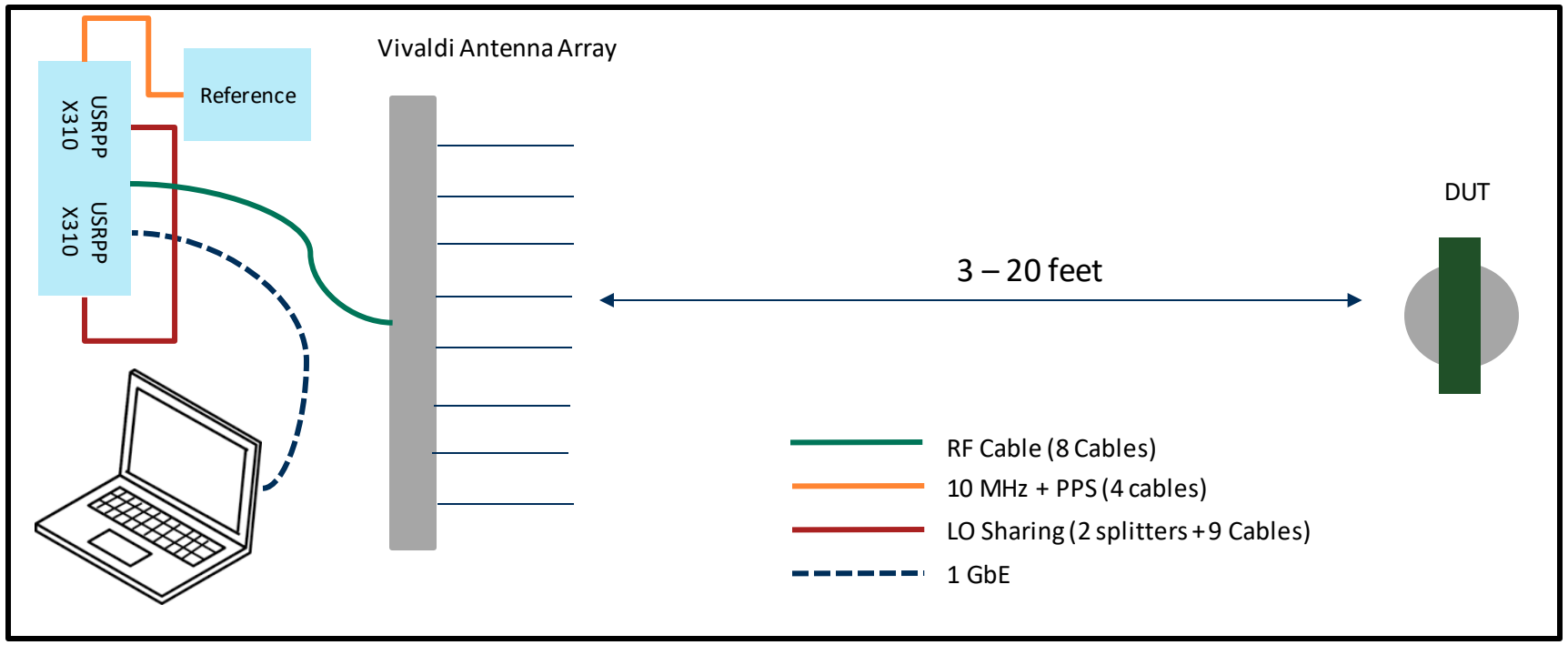
AEP Detects Change

Array Processing

Hardware Configuration

@1400 MHz, λ\2 = 10.7cm
@1210 MHZ, λ\2 = 12.4cm
@915 MHz, λ\2 = 16.4cm
@606 MHz, λ\2 = 24.8cm

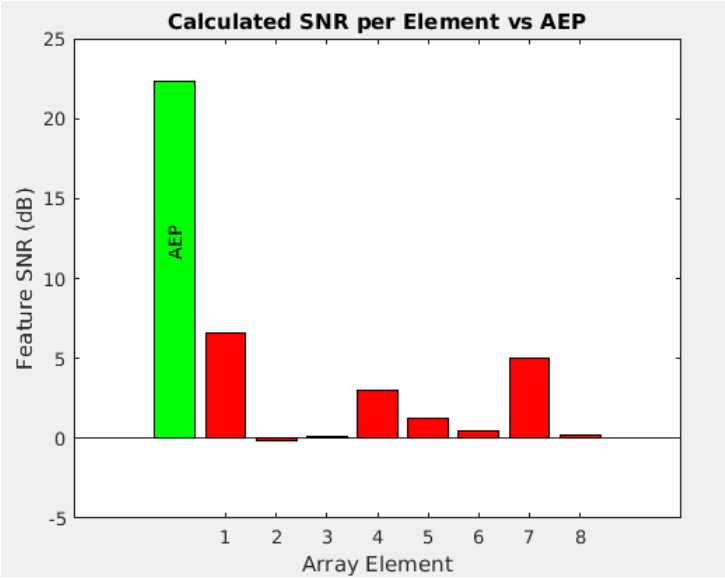
- Two Ettus USRP X310s with Dual TwinRX and externally shared LOs
- Compact array of Vivaldi slot antennas capturing various possible polarities that the device could be emanating

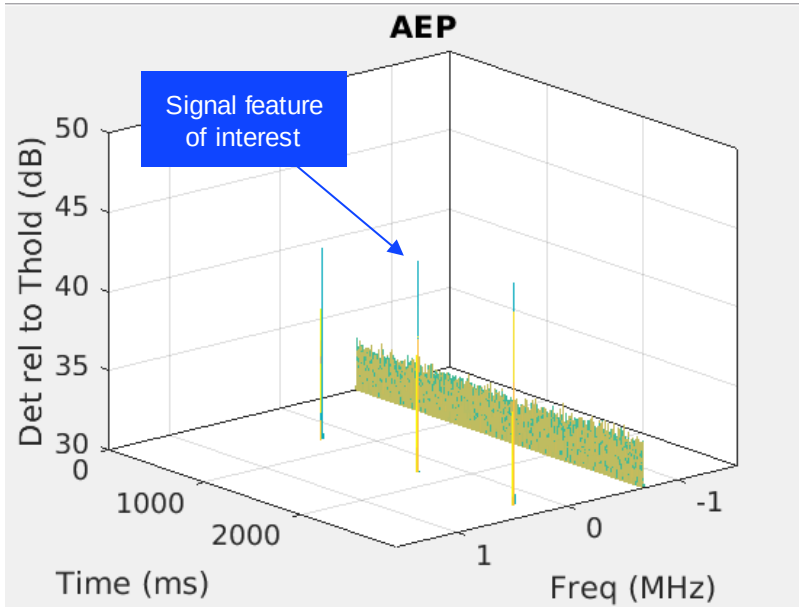


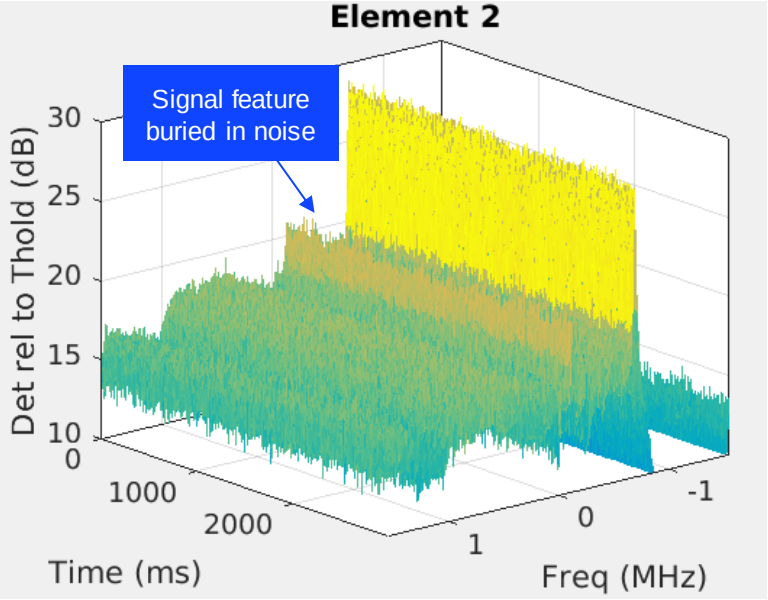
Array Processing

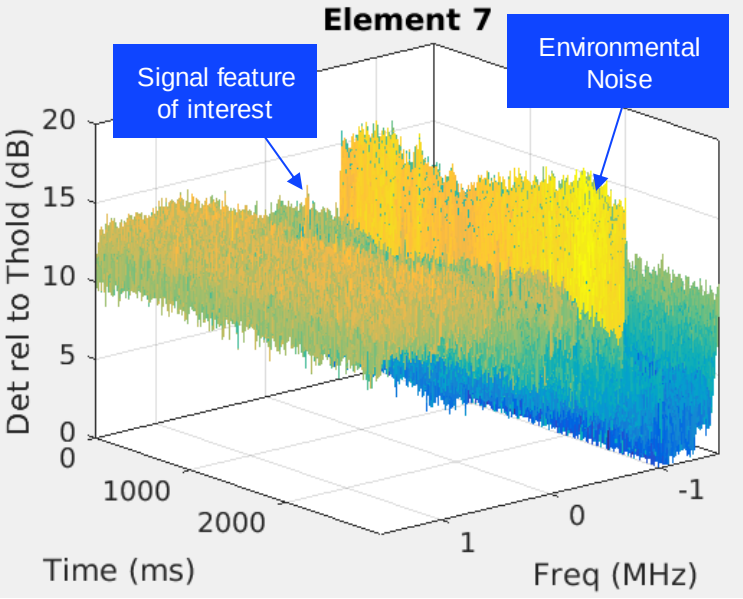
AEP Results

- Commercial device 20 ft away from antenna array
- Signal feature of interest in this case shows up on a particular subcarrier within this bandwidth
- Show gain in SNR on that particular subcarrier (which was identified via AEP metric)









Conclusions and Future Directions

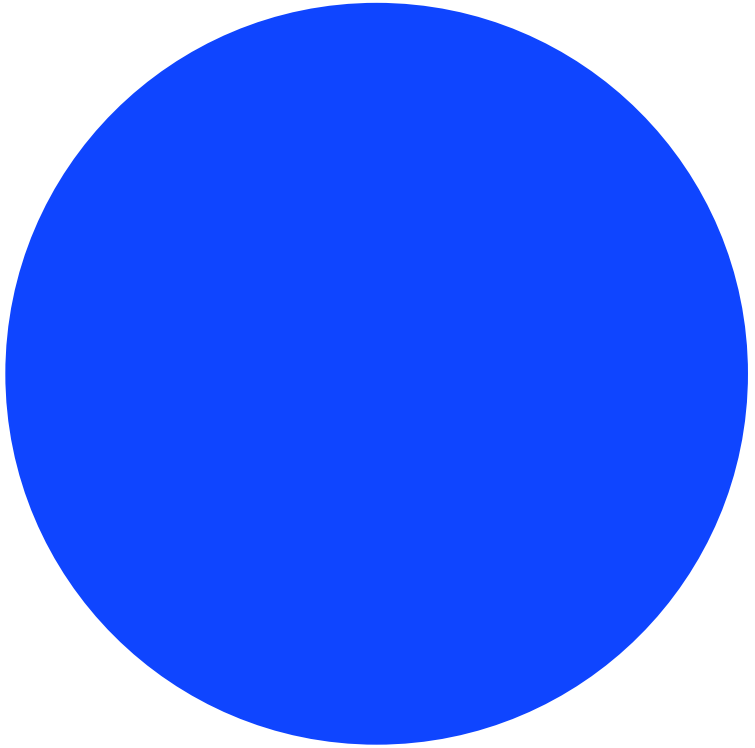
Conclusions

- For a wide range of devices, we can distinguish known/unknown code
  - Combination of technologies: signal processing, machine learning, program analysis, anomaly detection
- Gnuradio invaluable component in the CASPER system
  - ZMQ integration critical part of streaming system
- Efficient techniques to run at real-time speeds
- Commodity hardware for minimized deployment costs

Future Directions

- More robust feature extraction
- Power Modality
- Real-time AEP processing integrated into streaming system

Thank you



perspecta
LABS