Machine learning for wireless spectrum awareness

Enrico Mattei
Greg Harrison
Expedition Technology, Inc.

Acknowledgement:
This research was developed with funding from the Defense Advanced Research Projects Agency (DARPA).

The team


Afsayh Saquib

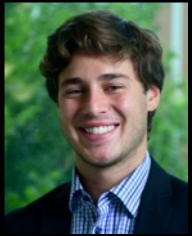

Brent Marin


Andrew Draganov


Cass Dalton


Enrico Mattei


Greg Harrison


Harry Engoren

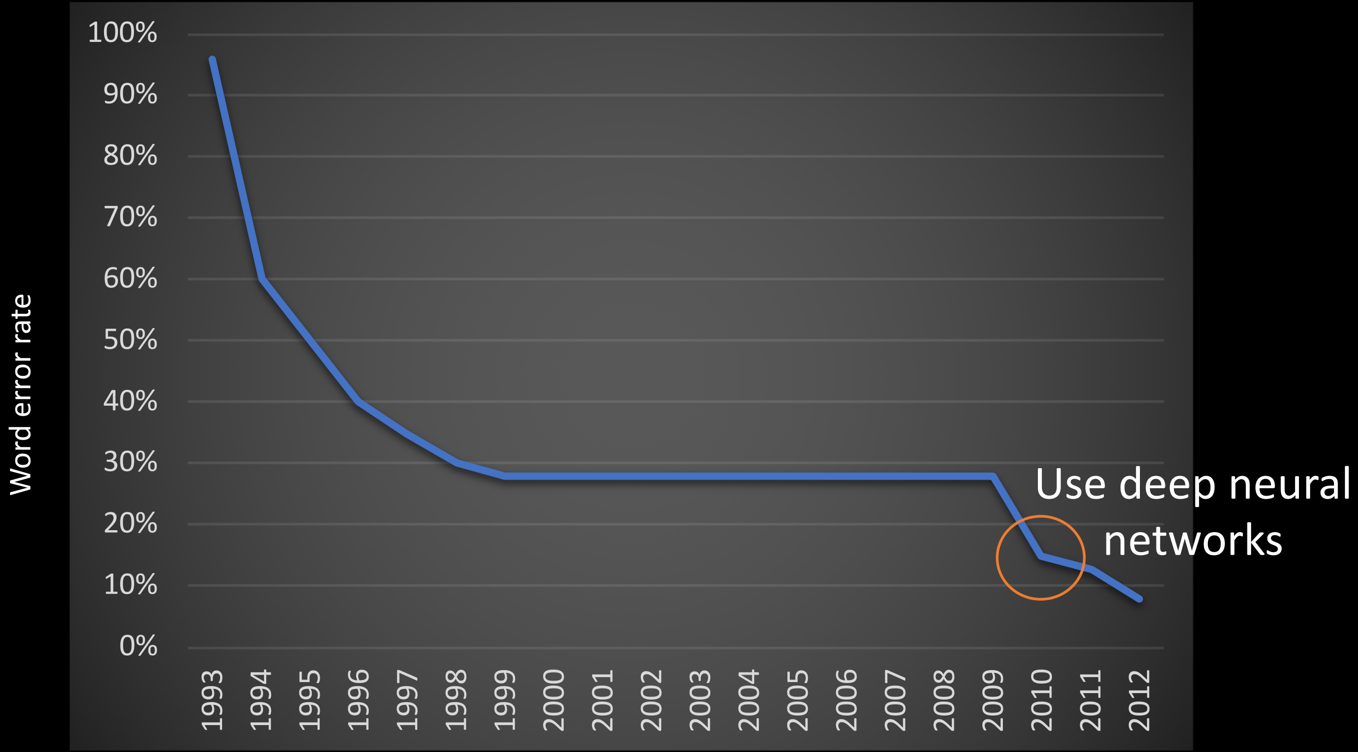

Mark Wallace


Ryan Crawford


Stella Yu

EXPEDITION TECHNOLOGY

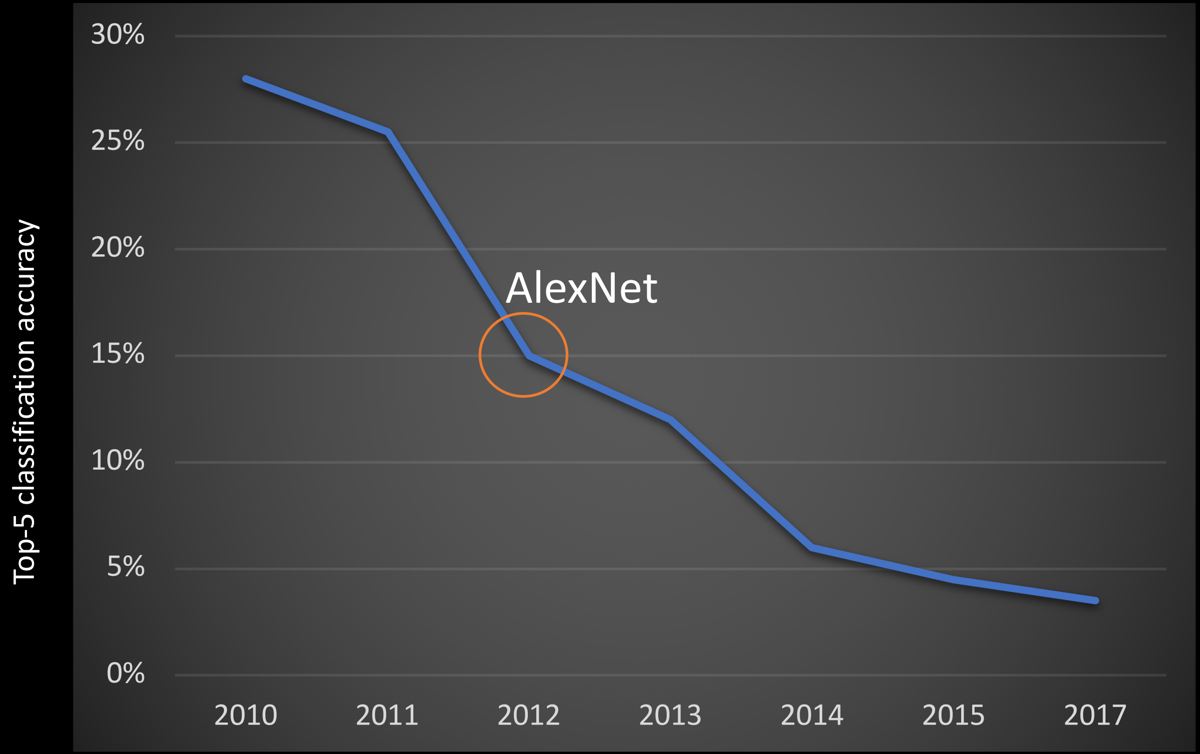
Spontaneous speech recognition



Source data:NIST, Microsoft

Approved for Public Release, Distribution Unlimited

Image recognition



Source data: image-net.org

The RF community is catching up!

Can we have simpler, machine learned systems?

- This is the era of AI and big data!
- The RF/wireless community is catching up
- Numerous applications:
    - Modulation/demodulation
    - Equalization
    - Channel coding
    - Modulation recognition …
    - … and more

https://www.comsoc.org/jsac/cfp/machine-learning-wireless-communication

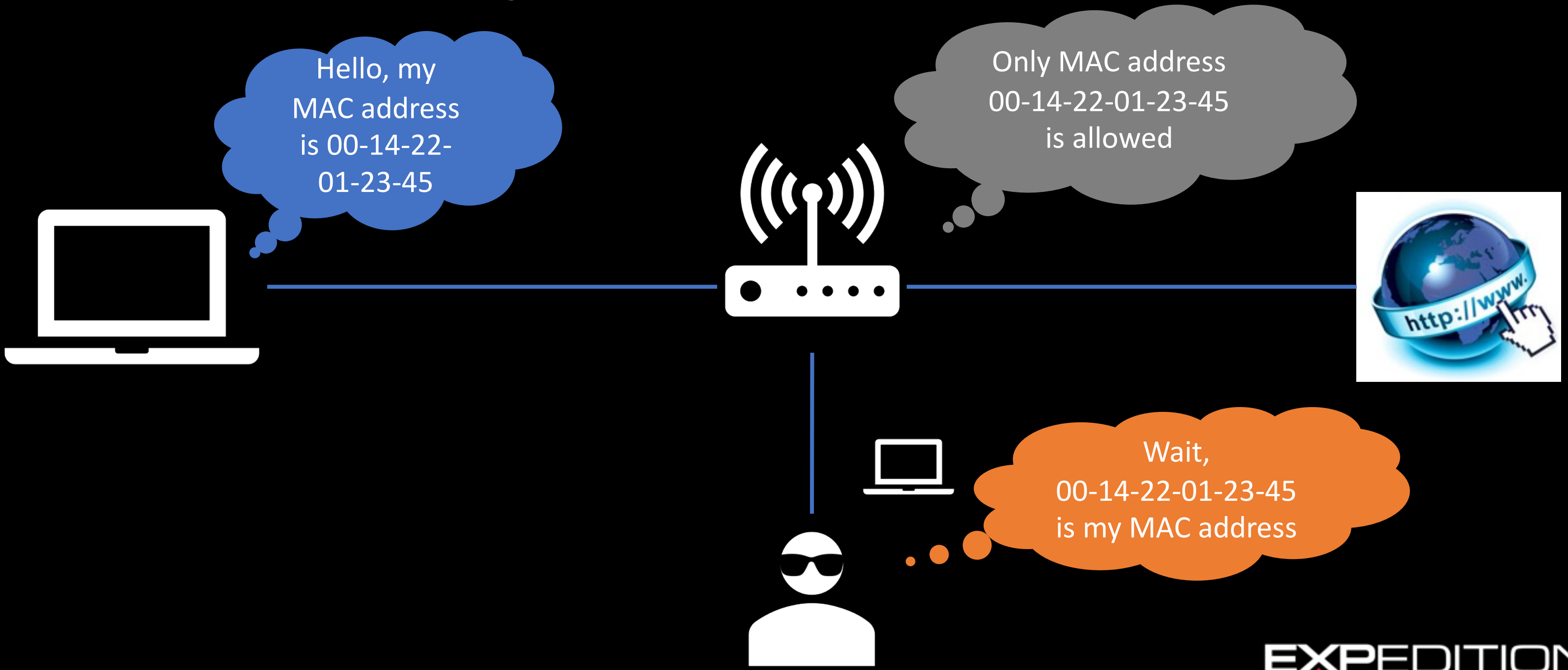
Tim O'Shea and Jakob Hoydis, *"An introduction to deep learning for the physical layer,"* IEEE Trans. On Cognitive Comms and Networking, Vol. 3 Issue 4, Dec 2017

Radio Frequency Machine Learning Systems (DARPA-RFMLS)

Task 1: Security in the internet of things (IoT)

IoT security

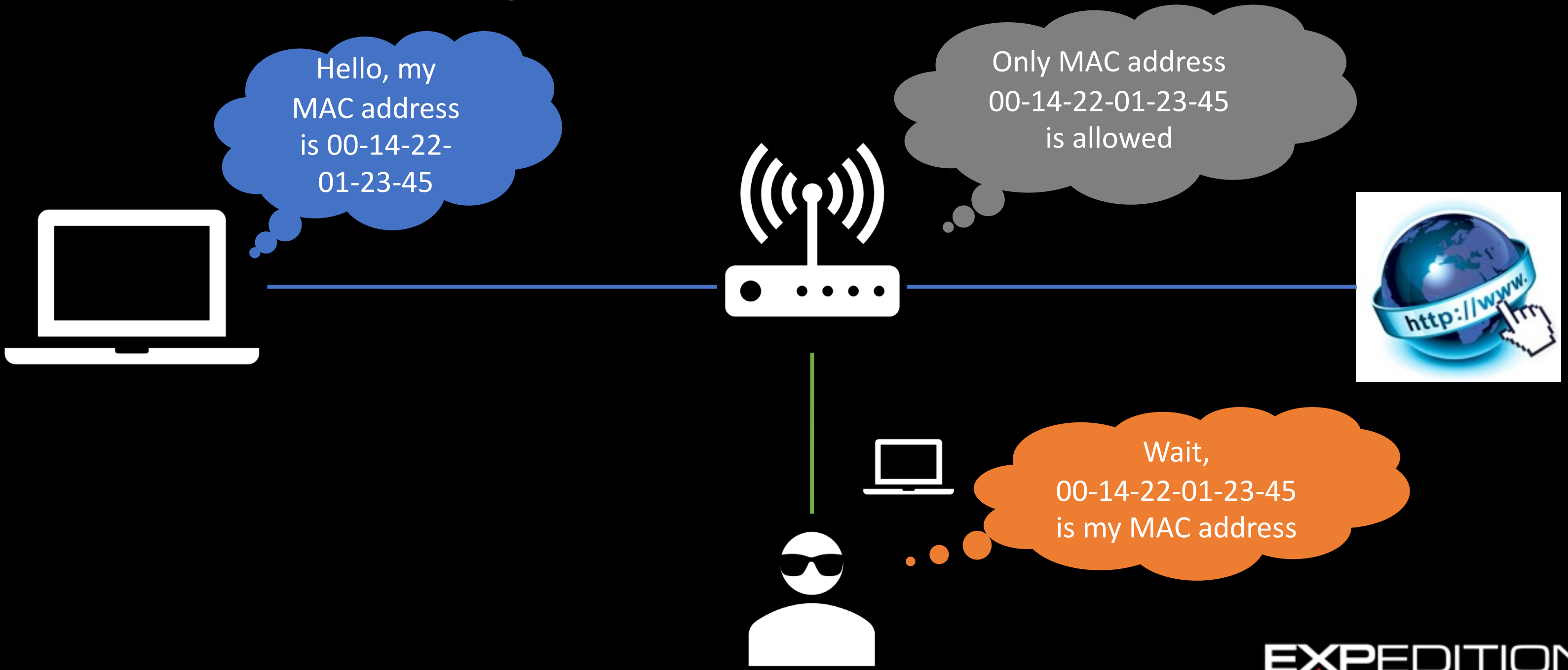
- MAC addresses are easily spoofed
- Nefarious actor could get unauthorized access



EXPEDITION TECHNOLOGY

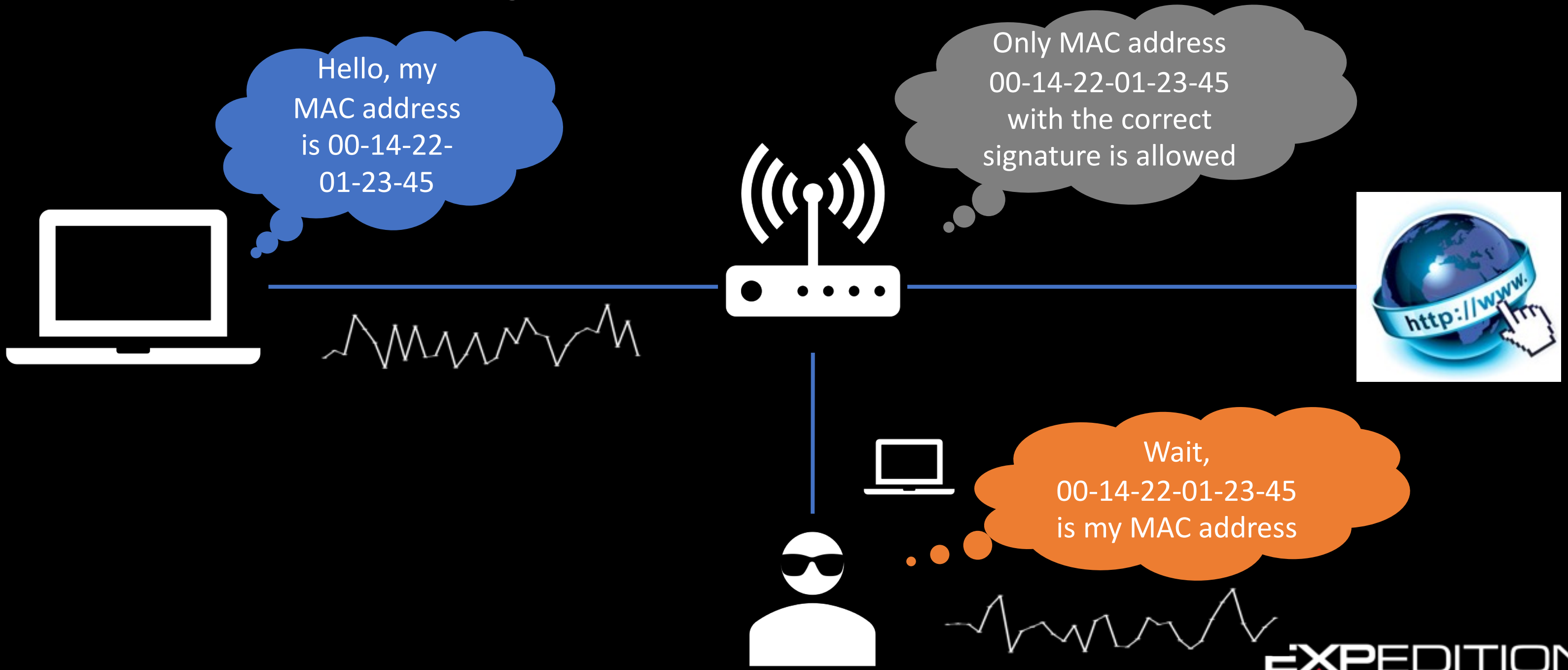
IoT security

- MAC addresses are easily spoofed
- Nefarious actor could get unauthorized access



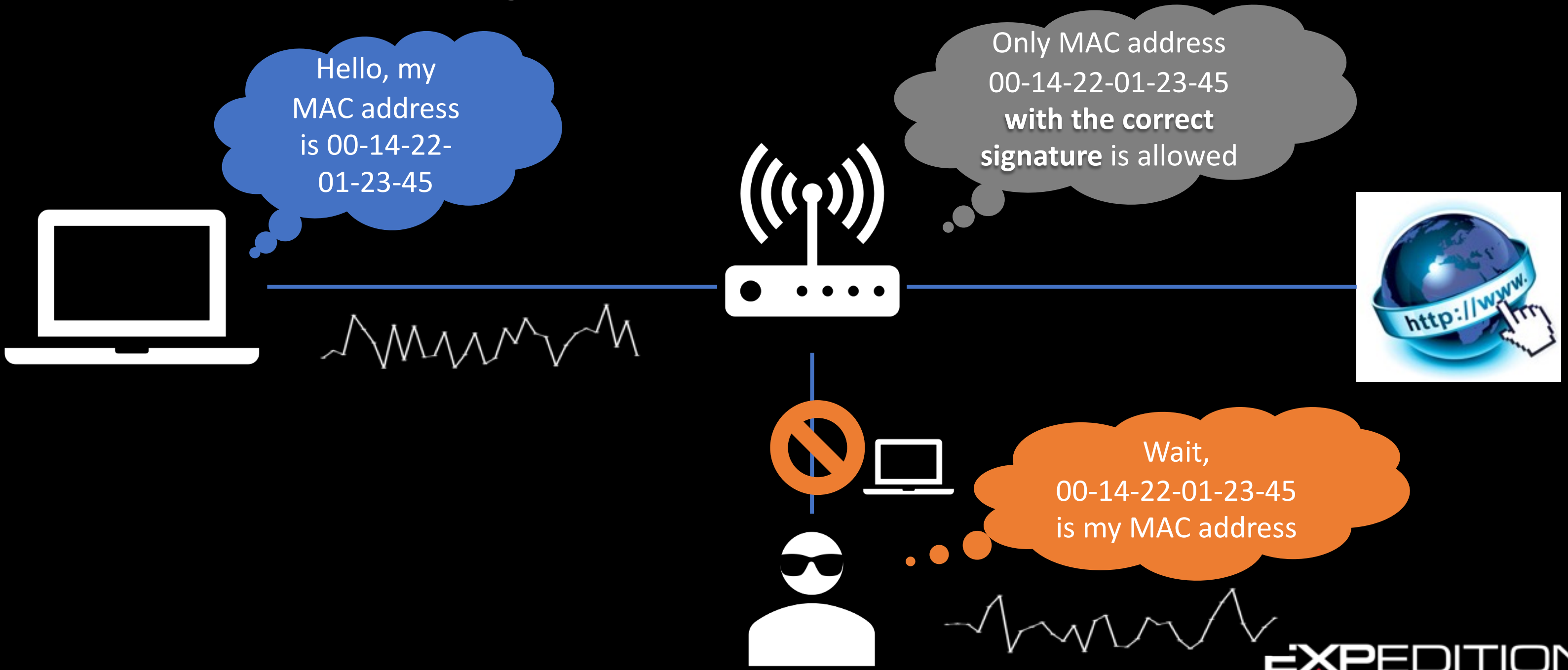
IoT security: signal characterization

- MAC addresses are easily spoofed
- Nefarious actor could get unauthorized access



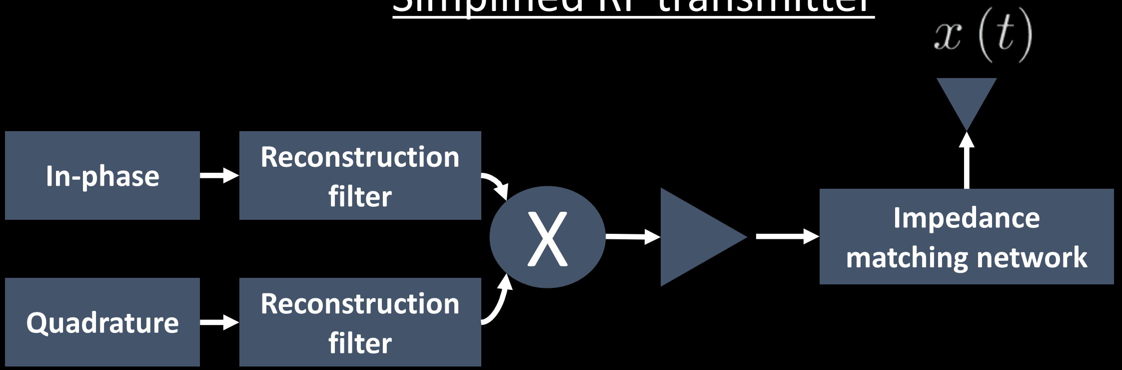
IoT security: signal characterization

- MAC addresses are easily spoofed
- Nefarious actor could get unauthorized access

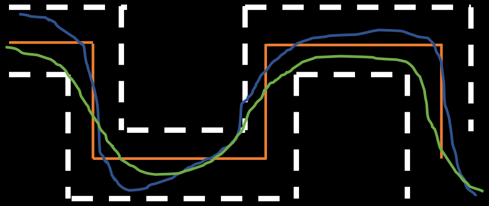


IoT security: signal characterization

Simplified RF transmitter



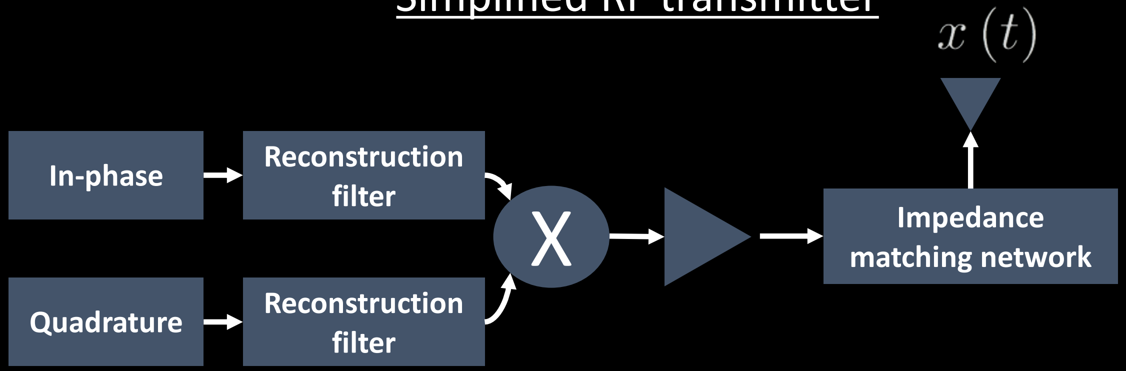
Received bit transitions



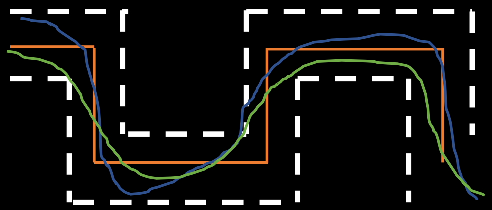
$$x(t) = Re\left[\left(\left(h_{re} * Re\left[a(t)\, e^{j\omega(t)}\right]\right)(t) + j\left(h_{im} * Im\left[a(t)\, e^{j\omega(t)}\right]\right)(t)\right) e^{j2\pi f_c t}\right]$$

IoT security: signal characterization

Simplified RF transmitter



Received bit transitions

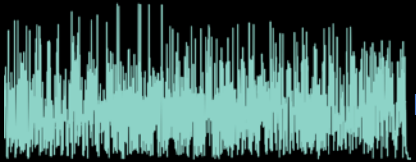


Signal characterization can be achieved by measuring the relationship between the in-phase and quadrature impulse responses of the reconstruction filters

IoT security: signal characterization

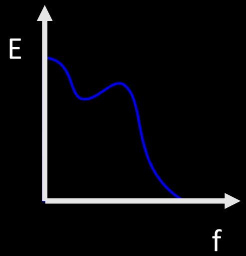
Prior art looks to extract discriminative features by hand:

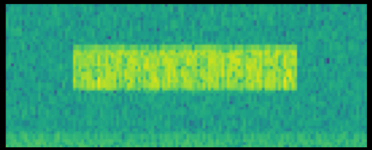
- Signal/modulation specific
- Need full signal demodulation
- Time consuming and involved process
- Usually designed to work with the dataset at hand (no generalization)

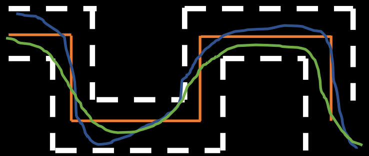


Input signal (complex IQ)

Detection and carrier recovery → Demodulation → Characterize the “wiggles”







IoT security: feature learning

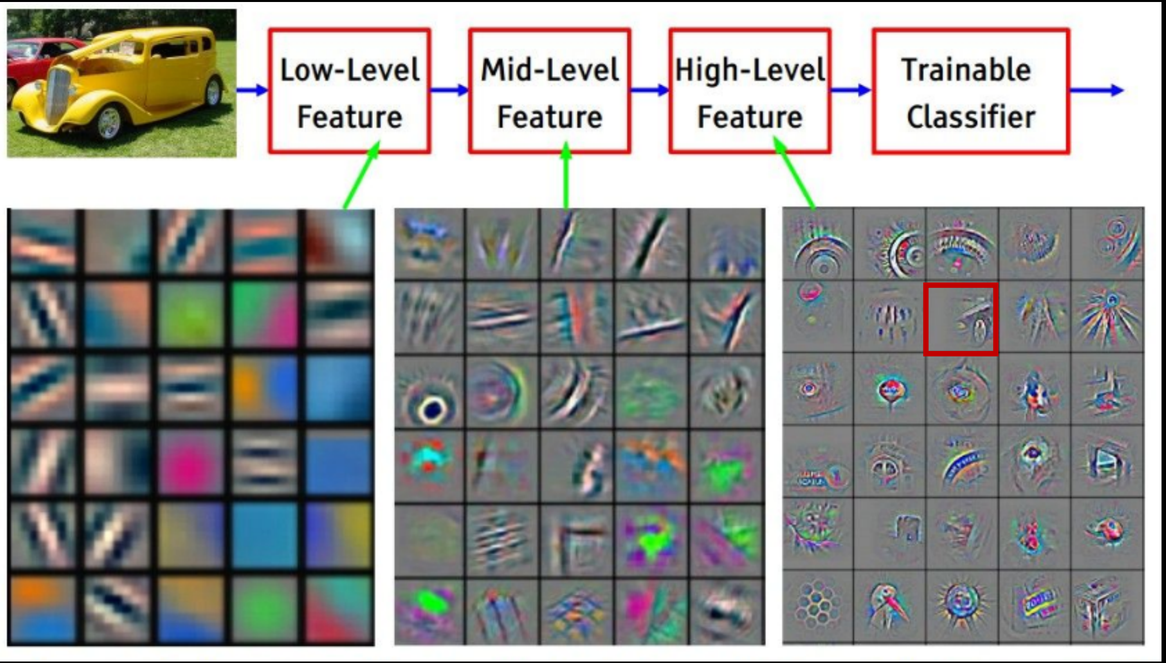
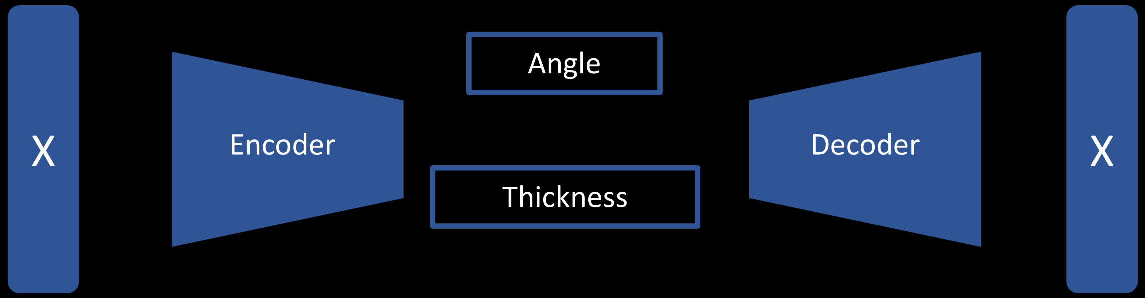


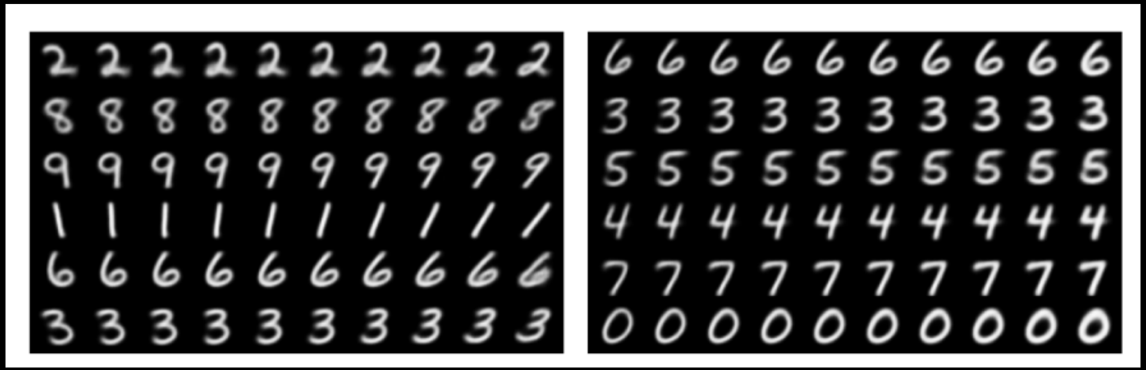
Image courtesy of Y. LeCun

EXPEDITION TECHNOLOGY

Approved for Public Release, Distribution Unlimited

IoT security: feature learning

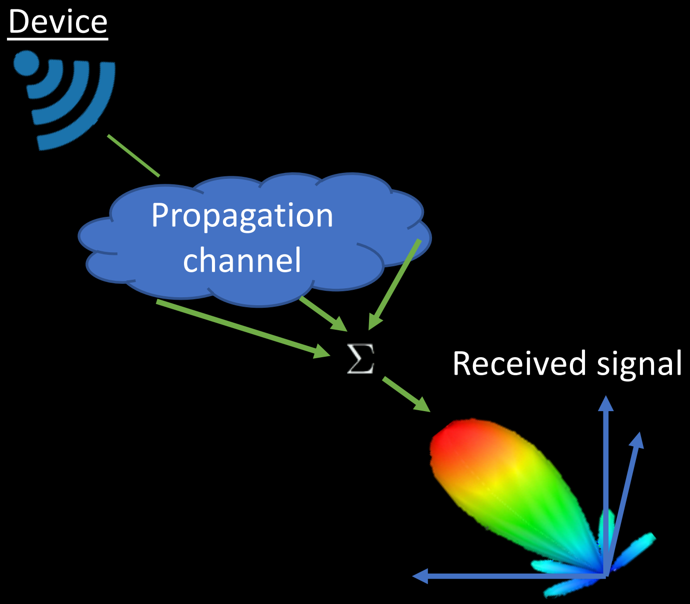


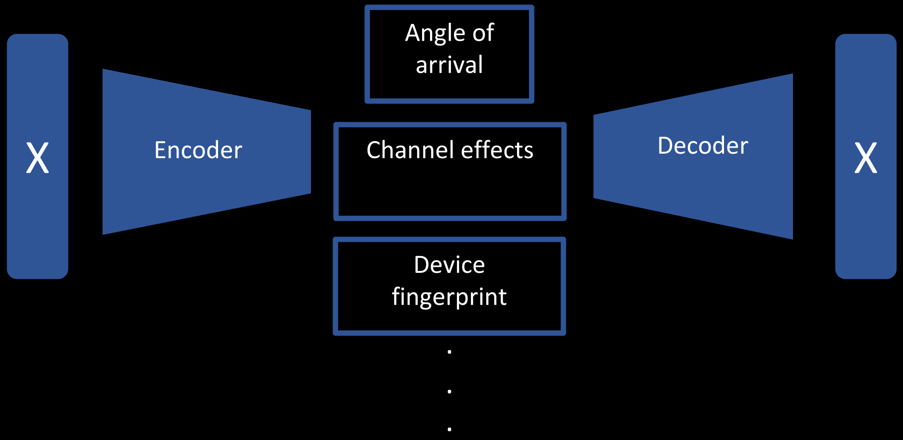


Angle Thickness

IoT security: feature learning

Structured representations to address the challenges in the wireless domain





IoT security: feature learning

A generative signal model:

$$y \sim cat\left(1/K\right)$$

Device label

$$\mathbf{z}|y \sim \mathcal{CN}\left(\mu_{\mathbf{z}}\left(y\right), diag\left(\sigma_z\left(y\right)\right)\right)$$

Latent signature

$$\mathbf{x}|\mathbf{z} \sim \mathcal{CN}\left(\mu_{\mathbf{x}}\left(\mathbf{z}\right), diag\left(\sigma_x\left(\mathbf{z}\right)\right)\right)$$

Observed data

$$ELBO = \mathbb{E}_{q(\mathbf{z}|\mathbf{x})}\left[log\, \underbrace{p\left(\mathbf{x}|\mathbf{z}\right)}_{\text{Decoder}}\right] - KL\left(\underbrace{q\left(\mathbf{z}|\mathbf{x}\right)}_{\text{Encoder}} \| \underbrace{p\left(\mathbf{z}|y\right)}_{\text{Prior}}\right)$$

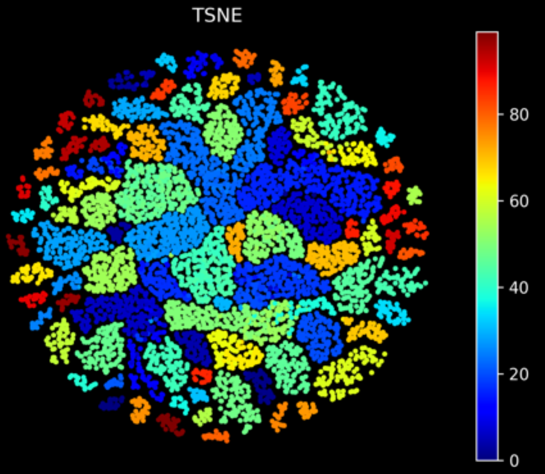
IoT security: feature learning

A generative signal model:

$$y \sim cat\left(1/K\right)$$ Device label

$$\mathbf{z}|y \sim \mathcal{CN}\left(\mu_{\mathbf{z}}\left(y\right), diag\left(\sigma_z\left(y\right)\right)\right)$$ Latent signature

$$\mathbf{x}|\mathbf{z} \sim \mathcal{CN}\left(\mu_{\mathbf{x}}\left(\mathbf{z}\right), diag\left(\sigma_x\left(\mathbf{z}\right)\right)\right)$$ Observed data



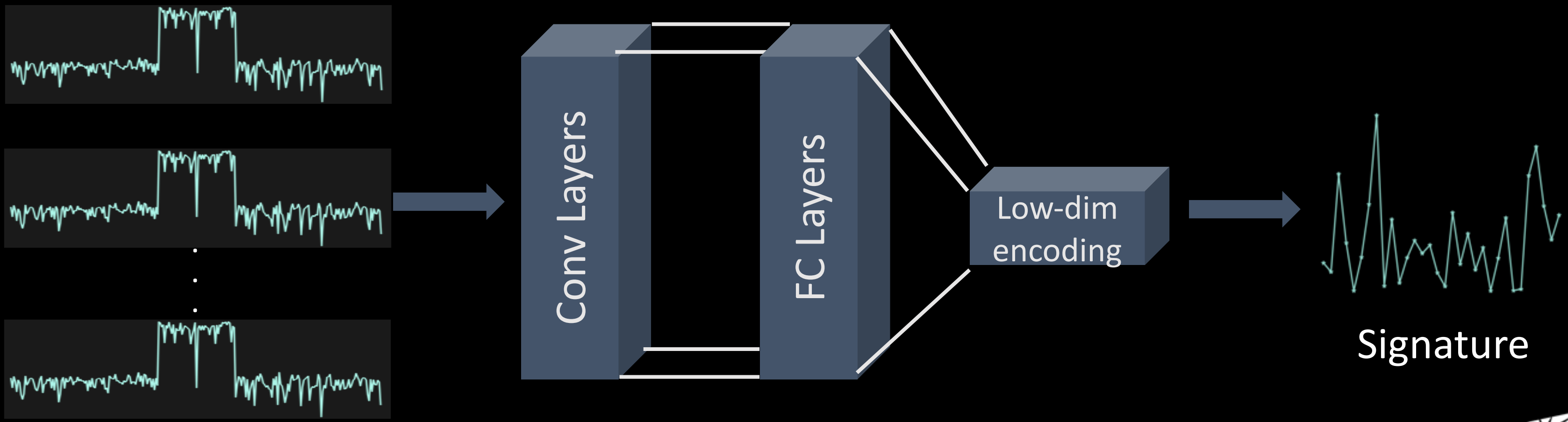
Source: EXP

In this model, the marginal distribution of the latent signature is a mixture of Gaussians

Approved for Public Release, Distribution Unlimited

IoT security: feature learning

$$\mathcal{L}\left(\theta, \phi; \boldsymbol{X}\right) = \sum_{i=1}^{M} \mathbb{E}_{q_\phi(\overline{\boldsymbol{z}}|\boldsymbol{X})}\left[\log p_\theta\left(\boldsymbol{x}_i|\overline{\boldsymbol{z}}\right)\right] - D_{KL}\left(\underbrace{q_\phi\left(\overline{\boldsymbol{z}}|\boldsymbol{X}\right)}_{\text{Encoder}} || \, p_\theta\left(\boldsymbol{z}|y\right)\right)$$

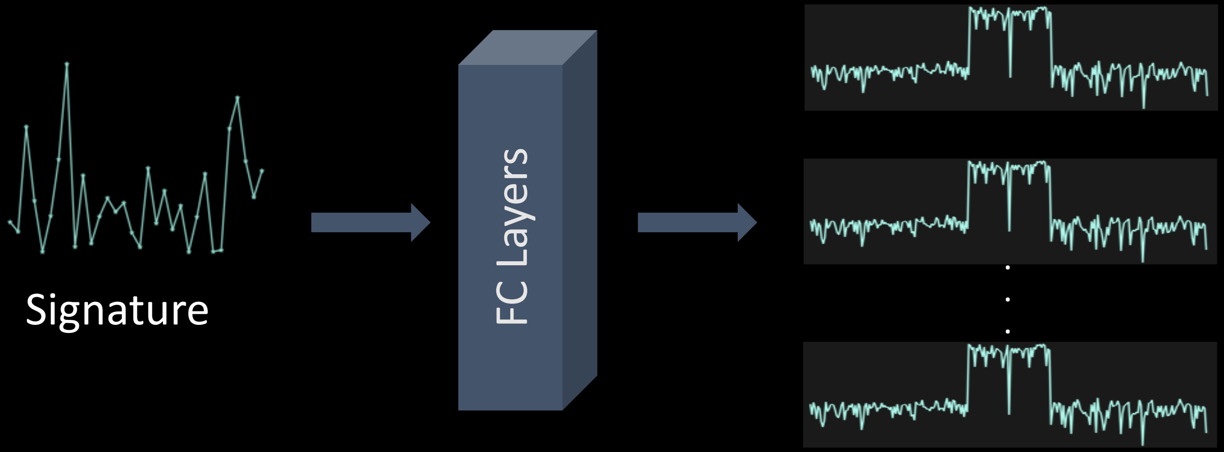


Sequence of DFT inputs

EXPEDITION TECHNOLOGY

IoT security: feature learning

$$\mathcal{L}\left(\theta, \phi; \boldsymbol{X}\right) = \sum_{i=1}^{M} \mathbb{E}_{q_\phi(\overline{\boldsymbol{z}}|\boldsymbol{X})}\left[\log \underbrace{p_\theta\left(\boldsymbol{x}_i|\overline{\boldsymbol{z}}\right)}_{\text{Decoder}}\right] - D_{KL}\left(q_\phi\left(\overline{\boldsymbol{z}}|\boldsymbol{X}\right) || \, p_\theta\left(\boldsymbol{z}|y\right)\right)$$



Signature

Sequence of DFT outputs

EXPEDITION TECHNOLOGY

RFMLS datasets

Signal libraries: IoT security

- Large dataset (~11TB) total
- WiFi (~103M) and ADS-B (~3.5M) signals
- Signal metadata in SigMF format (JSON)
- Each metadata file is associated with one binary file that contains the signal data
- Each metadata file can contain thousands of annotations

WiFi

ADS-B

Large and diverse dataset specifically collected for Radio ML. The first of its class!

GNURadio

https://github.com/gnuradio/SigMF

EXPEDITION TECHNOLOGY

Signal libraries: IoT security

General data collection information:

- Licensing
- Sampling rate
- etc.

Annotations:

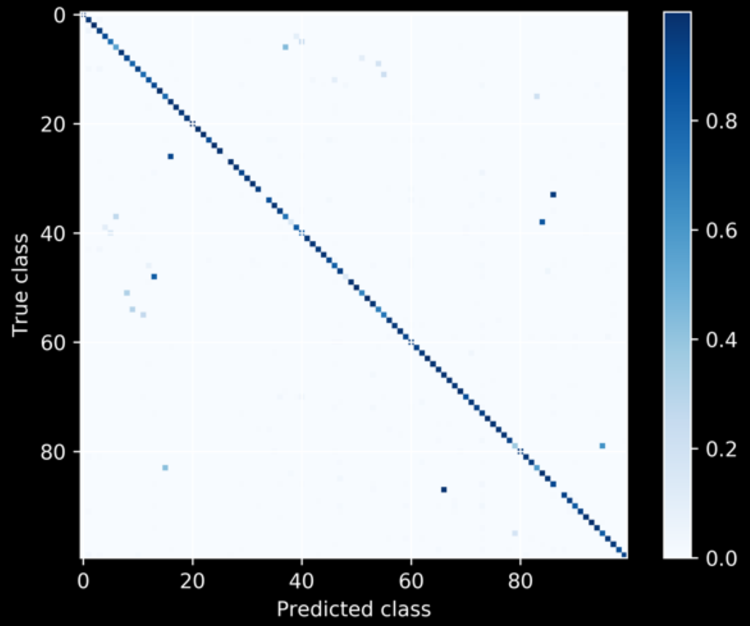
- Signal start sample
- Signal end sample
- label

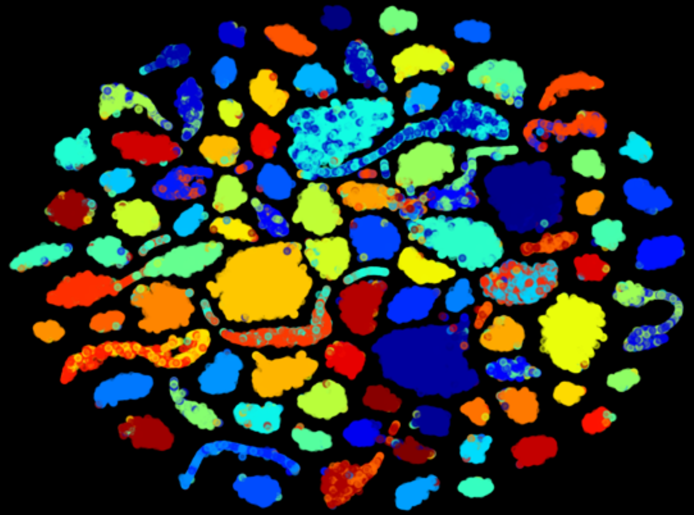

```
{
"global":{
"core:datatype": "ci16_le",
"core:sample_rate": 100000000.000000,
"core:version": "0.0.1",
"core:sha512":
"67bfe16fac96c6e70d3b56fab25e1813c35852cc847b33e9d812407fde5b8e7920f4ff5a67a479653860f83775810c6fe853f7b9baac3e
e5083ce51064185384",
"core:description": "Interleaved I&Q components.  A complete sample is 16 bits of I followed by 16 bits of Q.",
"core:license": "DARPA RFMLS data use agreement.",
"core:hw": "Receiver: TEKTRONIX,RSA5106B,B040879, FV:3.6.0239. LNA: ZEL-0812LN. Filter: None. Antenna: LP-1019
(vertical polarization).",
"core:extensions": {"capture_details":"required",
                    "rfml":"required"},
"capture_details:extension_version": "1.0",
"rfml:extension_version": "1.0",
"rfml:label_hierarchy": [["device_type"], ["device_id"]]
         },
"capture":[
 {
"core:sample_start": 0,
"core:frequency": 1090000000.000000,
"core:freq_lower_edge": 1089500000.000000,
"core:freq_upper_edge": 1090500000.000000,
"capture_details:acq_scale_factor": "9.313226e-13",
"capture_details:attenuation": 0,
"capture_details:acquisition_bandwidth": 80000000
 }
   ],
"annotations":[
 {
"core:sample_start": 3863880,
"core:sample_count": 6400,
"capture_details:SNRdB": 1.0,
"capture_details:signal_reference_number": "A-5561-10",
"rfml:label": ["ADS-B", "crane-gfi_3_dataset-4"]
 },
 {
"core:sample_start": 5696090,
"core:sample_count": 6400,
"capture_details:SNRdB": 2.8,
"capture_details:signal_reference_number": "A-5561-14",
"rfml:label": ["ADS-B", "crane-gfi_3_dataset-9"]
 },
```

Experimental results

Experimental results: 100 device identification

Sequence length = 1





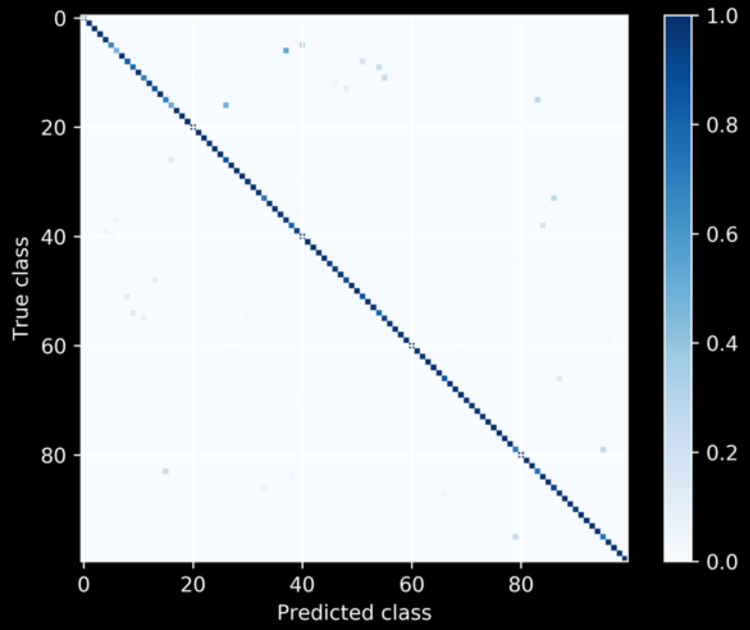
Source: EXP

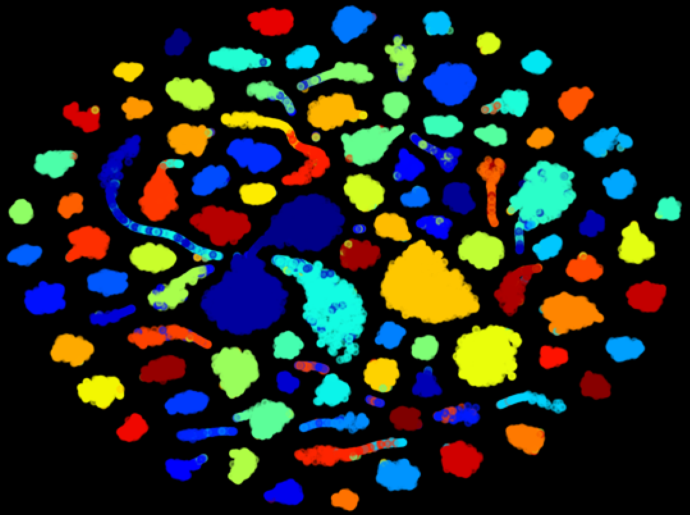
Approved for Public Release, Distribution Unlimited

EXPEDITION TECHNOLOGY

Experimental results: 100 device identification

Sequence length = 10



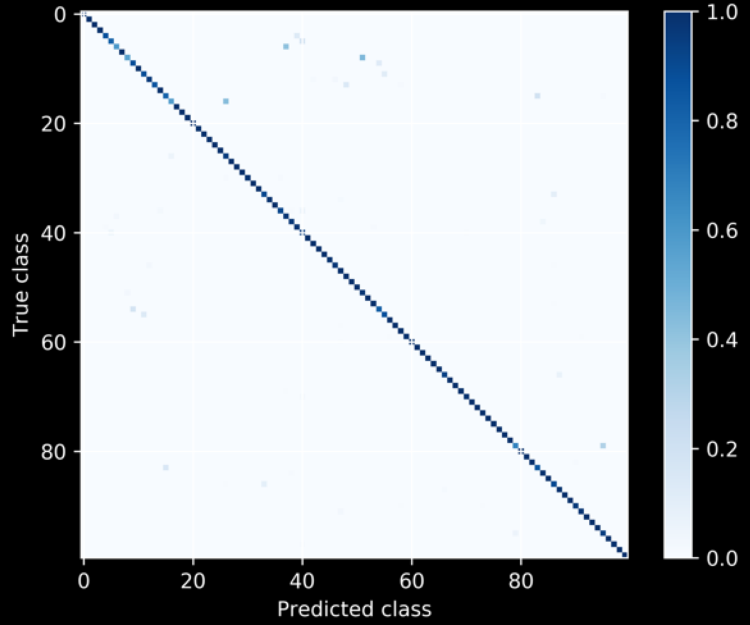


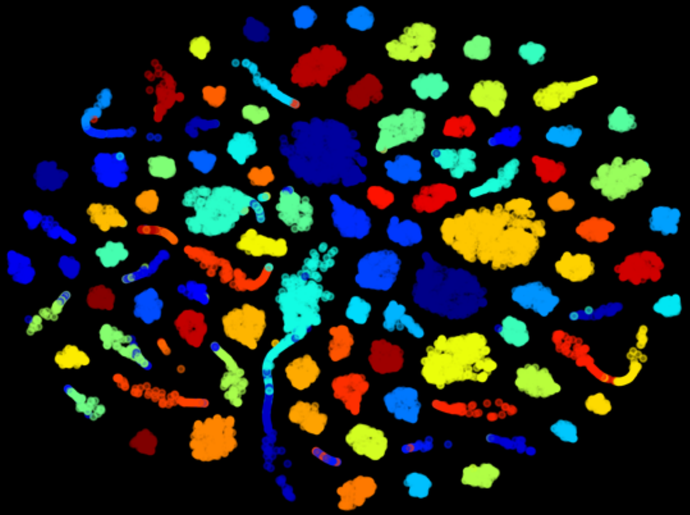
Source: EXP

Approved for Public Release, Distribution Unlimited

Experimental results: 100 device identification

Sequence length = 20





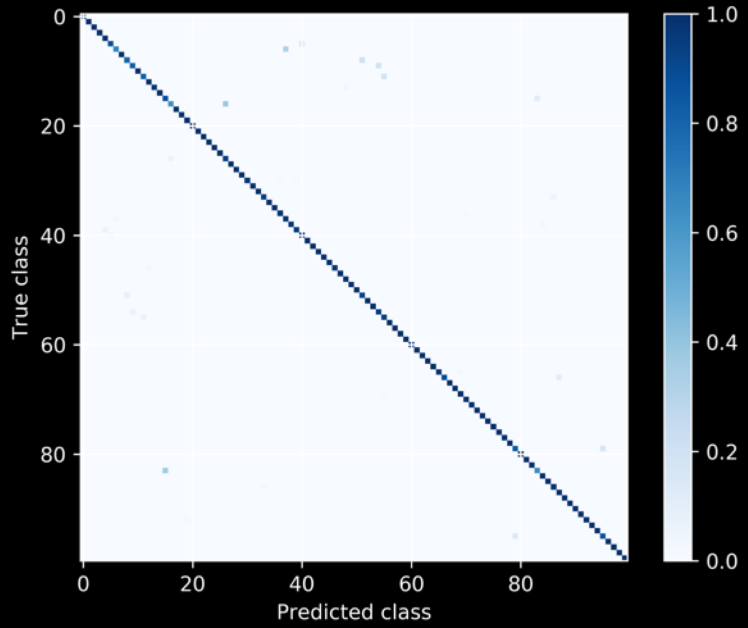
Source: EXP

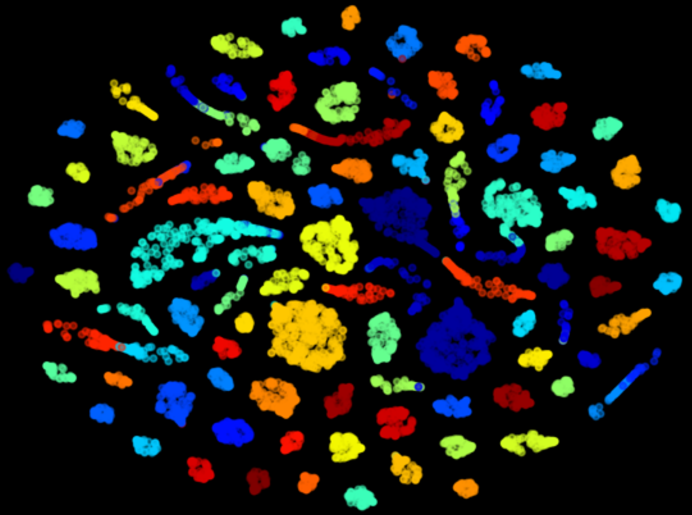
Approved for Public Release, Distribution Unlimited

EXPEDITION TECHNOLOGY

Experimental results: 100 device identification

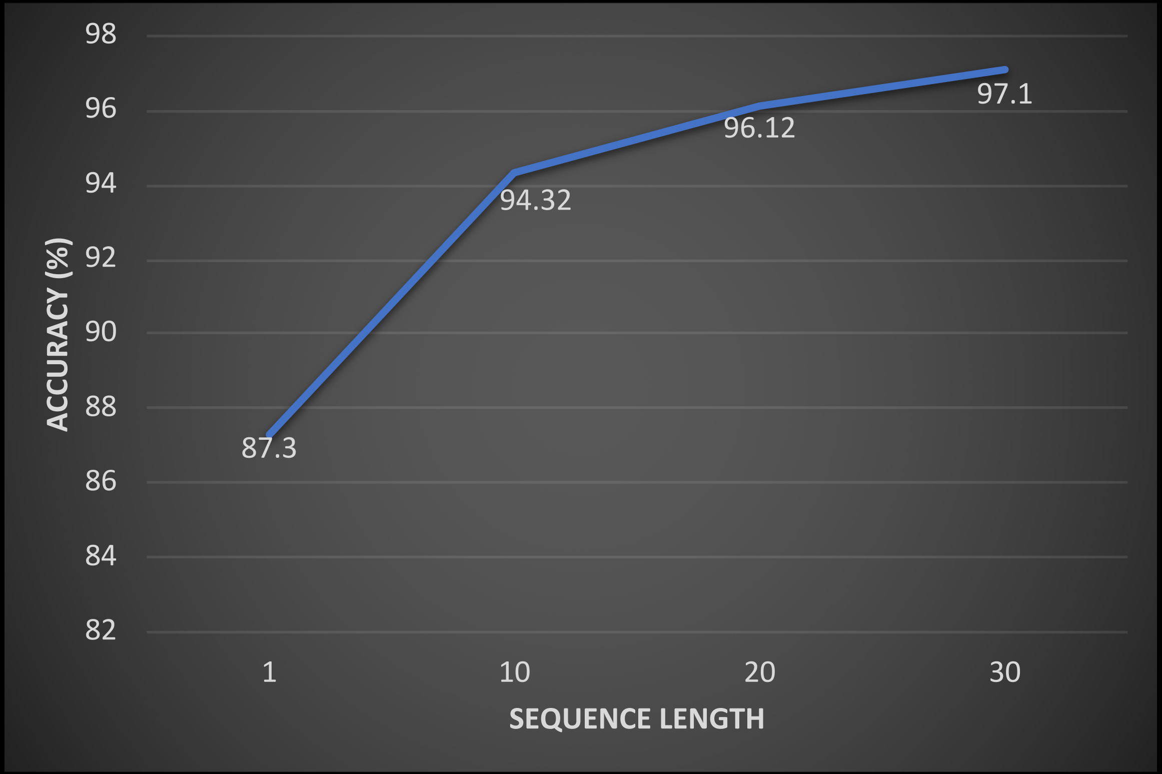
<u>Sequence length = 30</u>





Source: EXP

Experimental results: 100 device identification



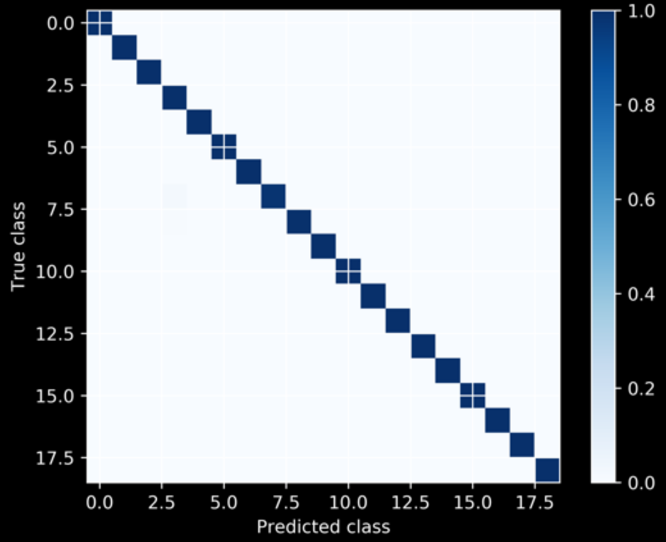
Source: EXP

Approved for Public Release, Distribution Unlimited

Experimental results: identification

19 bit-wise identical WiFi devices

- Verify that the machine learning system is not just recognizing the MAC address
- All devices were given the same MAC address

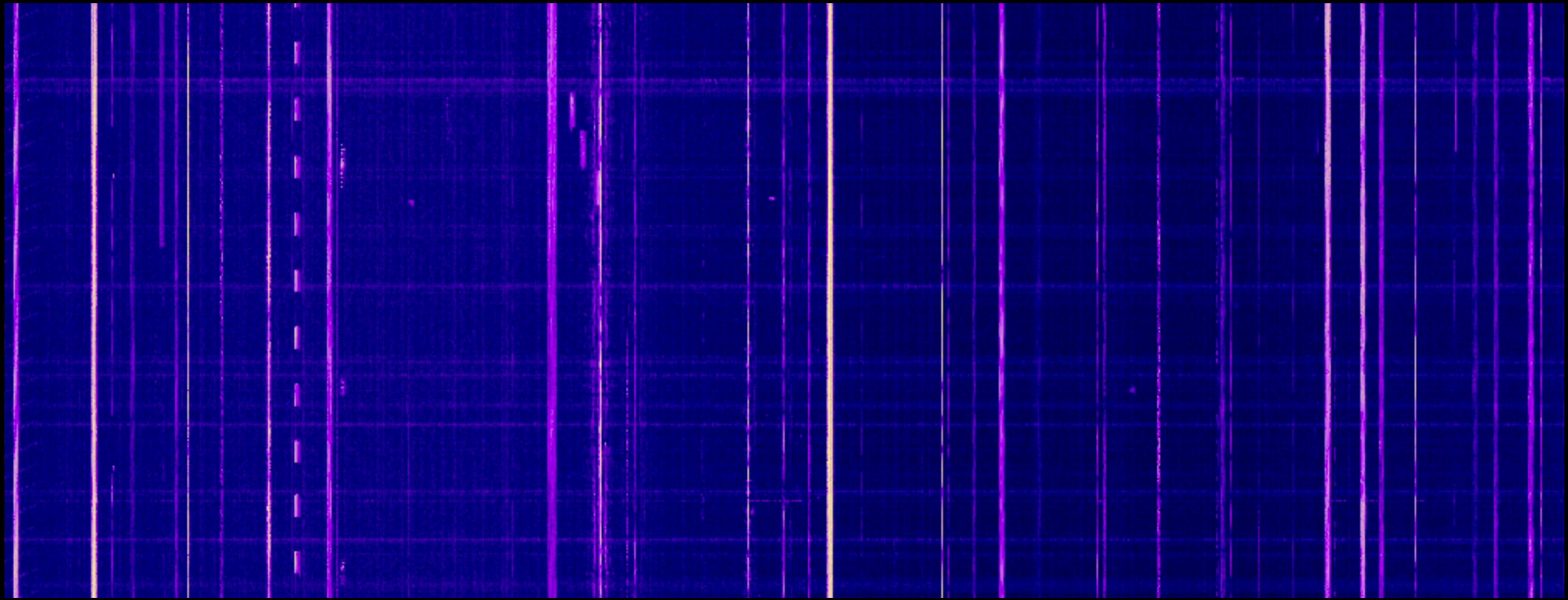


Source: EXP

Task 2: spectrum awareness — saliency and control

Spectrum awareness: saliency

Time



Frequency

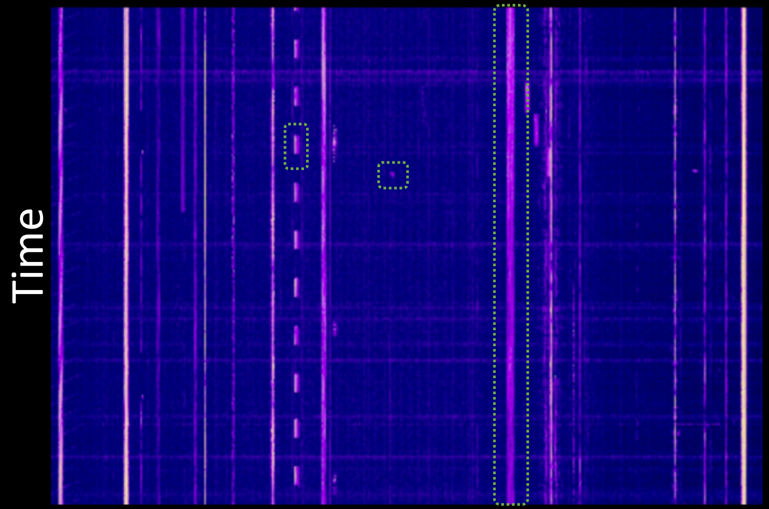
Video recorded using:
http://websdr.ewi.utwente.nl:8901/

Approved for Public Release, Distribution Unlimited

Spectrum awareness: saliency

- A wideband signal detection problem
- Traditional systems require “signal features” computed a-priori
- Systems are usually designed for the current target set
- Large signal variability in **time**, **frequency**, and **power**

WiFi is up to 800x wider than ADS-B. ADS-B is >40dB higher power than Bluetooth



http://websdr.ewi.utwente.nl:8901/

Spectrum awareness: saliency

The key question: Can we detect the ***important*** signals in wideband signal data?

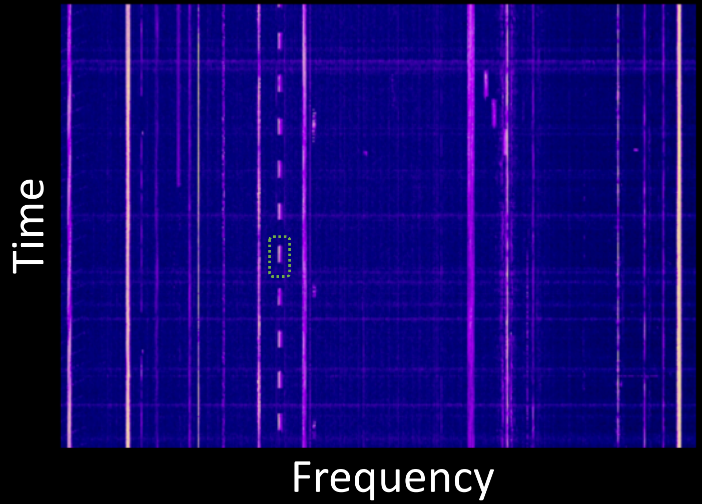
- ”Important” is defined by the user

Need to find the important signal in complex backgrounds

Example “important” signals are:

- APCO
- Bluetooth
- CDMA-2000 Forward
- CDMA-2000 Reverse
- DECT
- AIS

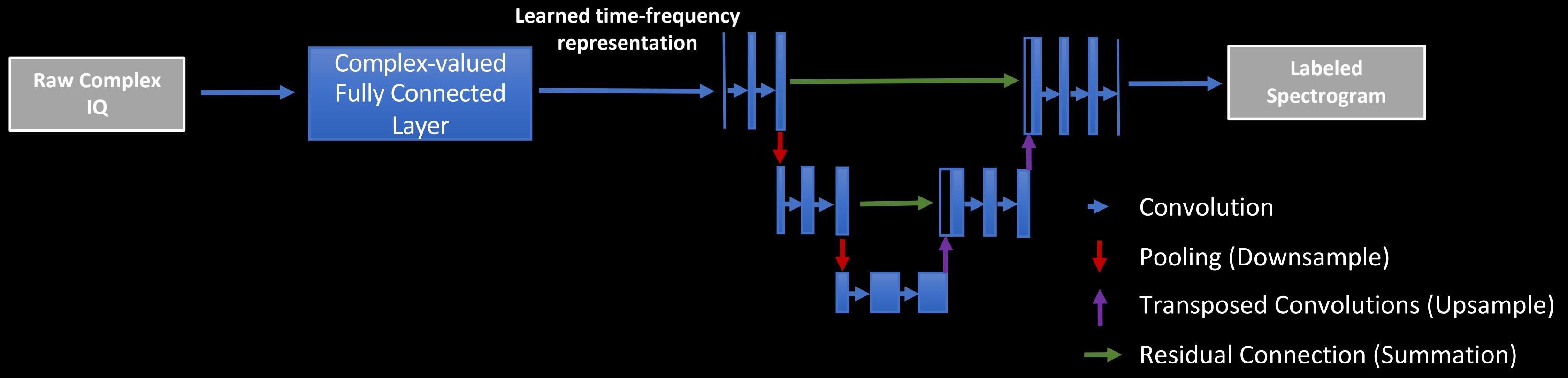
just to name a few



http://websdr.ewi.utwente.nl:8901/

Spectrum awareness: saliency

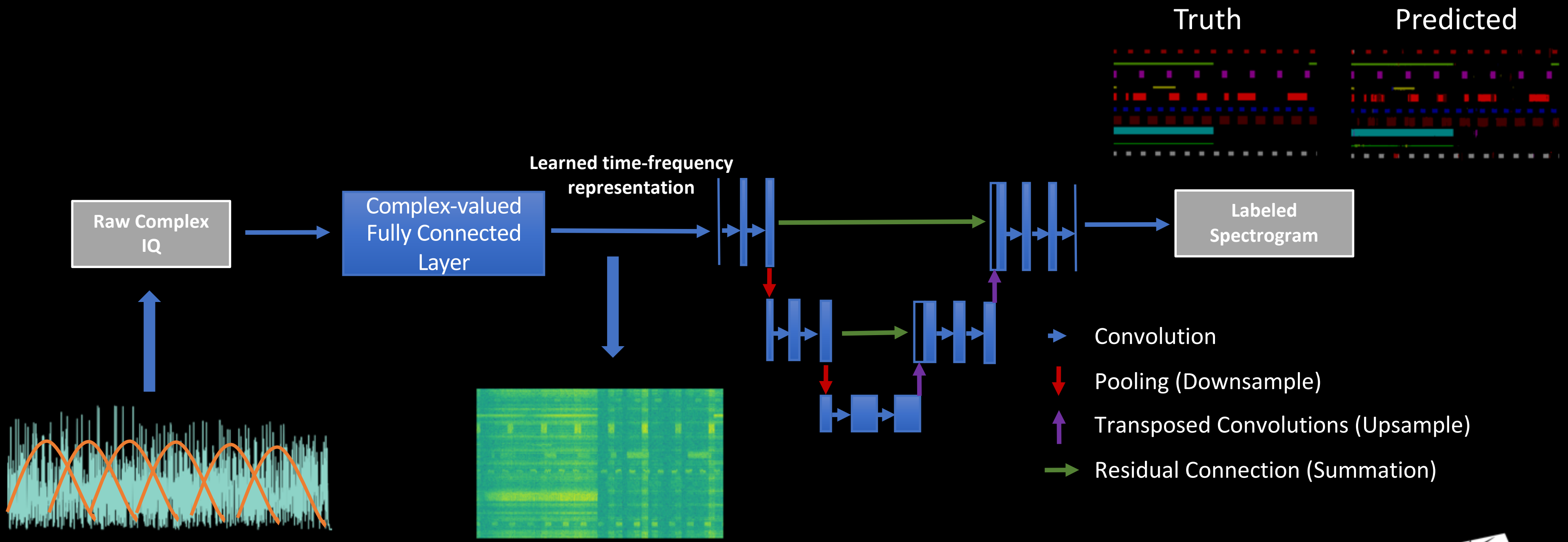
We can view this task as a semantic segmentation problem



Source: EXP

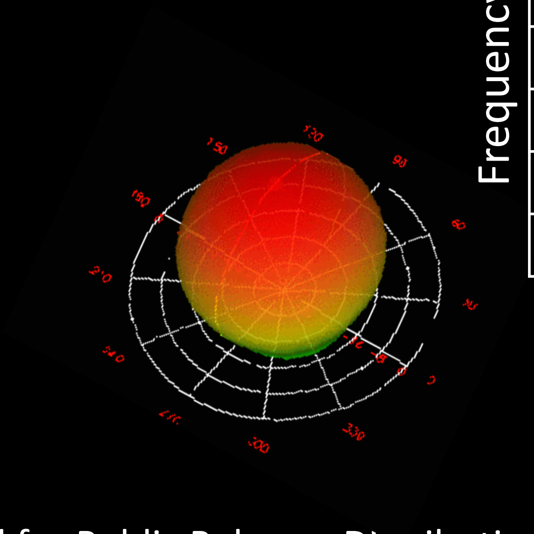
Spectrum awareness: saliency

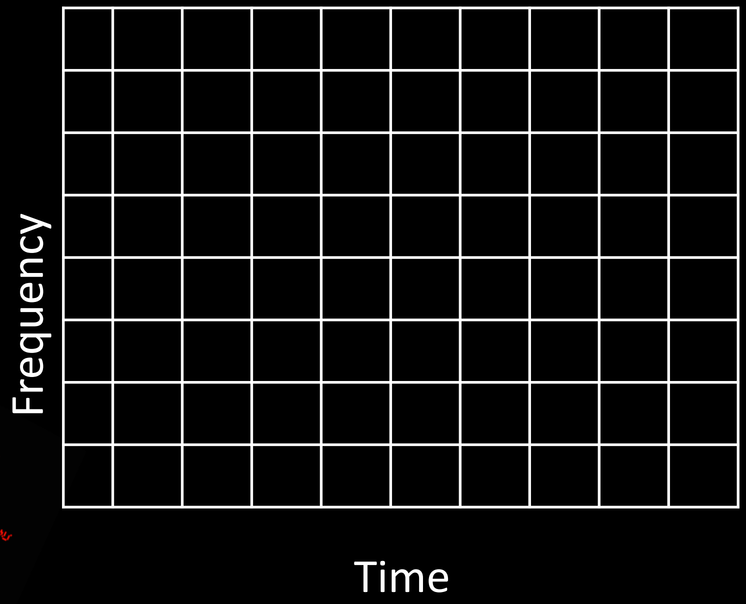
We can view this task as a semantic segmentation problem



Approved for Public Release, Distribution Unlimited

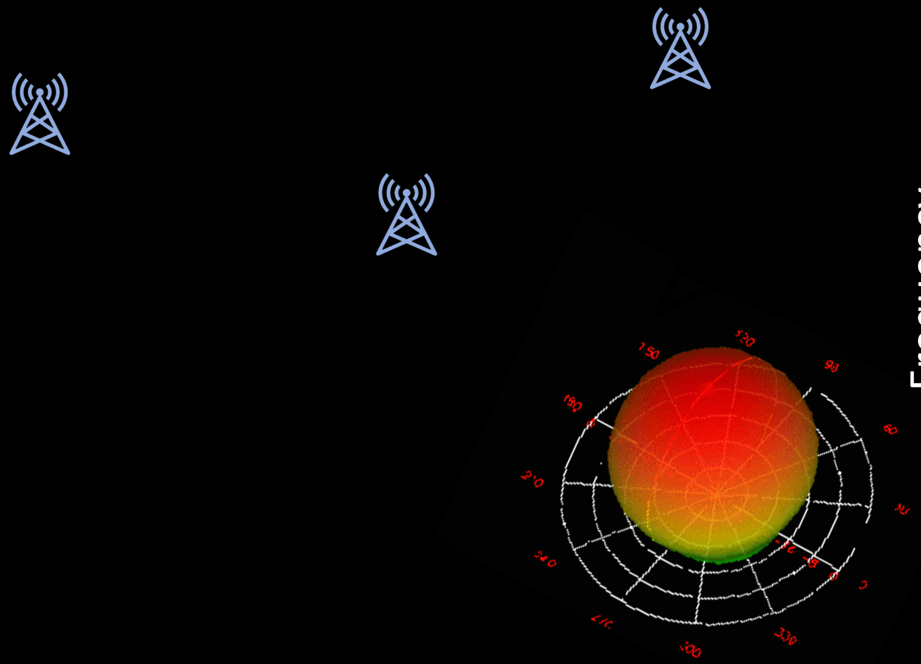
Spectrum awareness: control

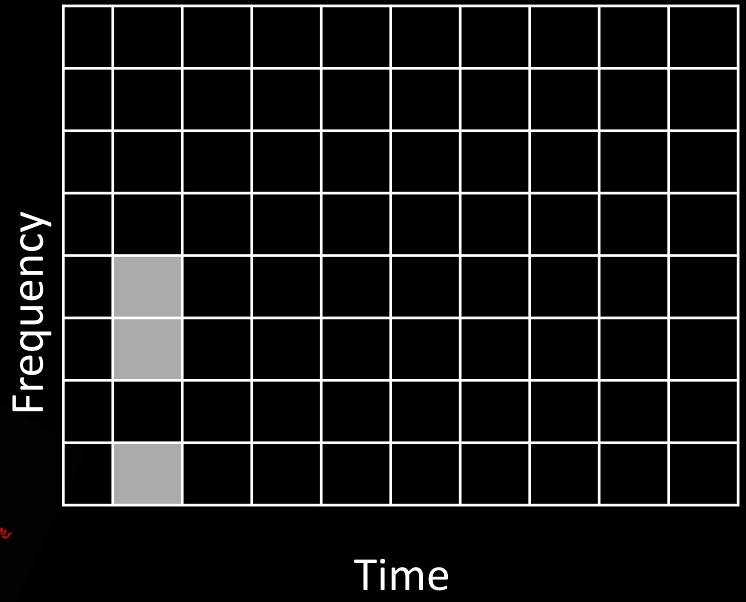




EXPEDITION TECHNOLOGY

Spectrum awareness: control

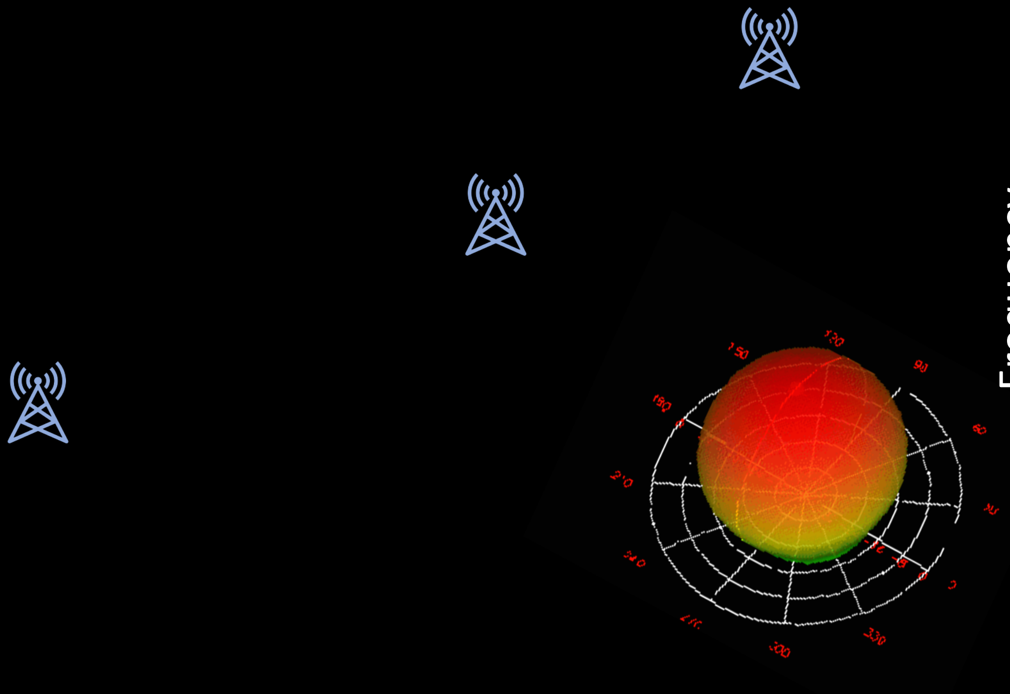


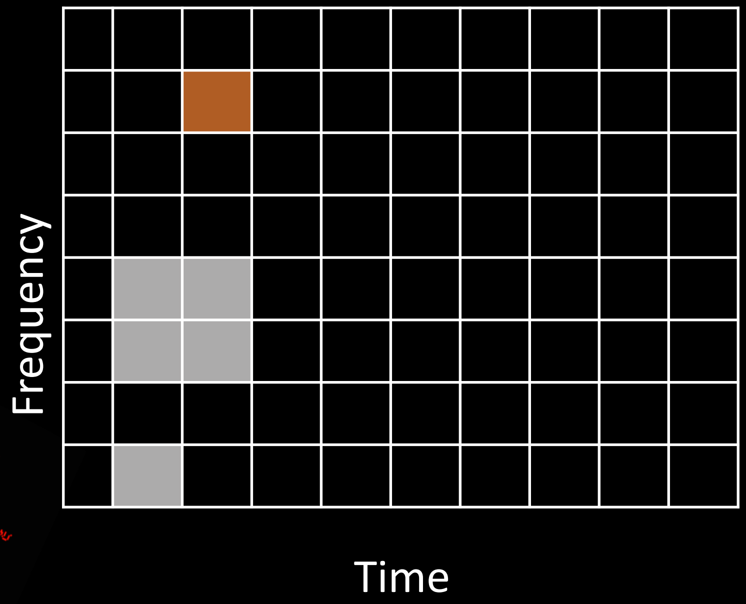


EXPEDITION TECHNOLOGY

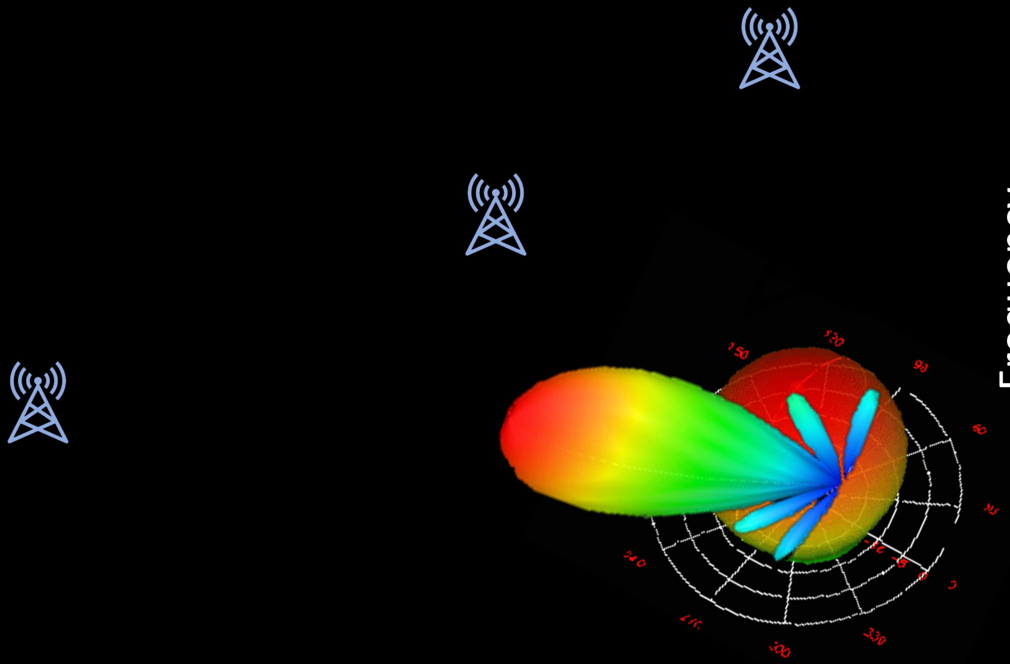
Approved for Public Release, Distribution Unlimited

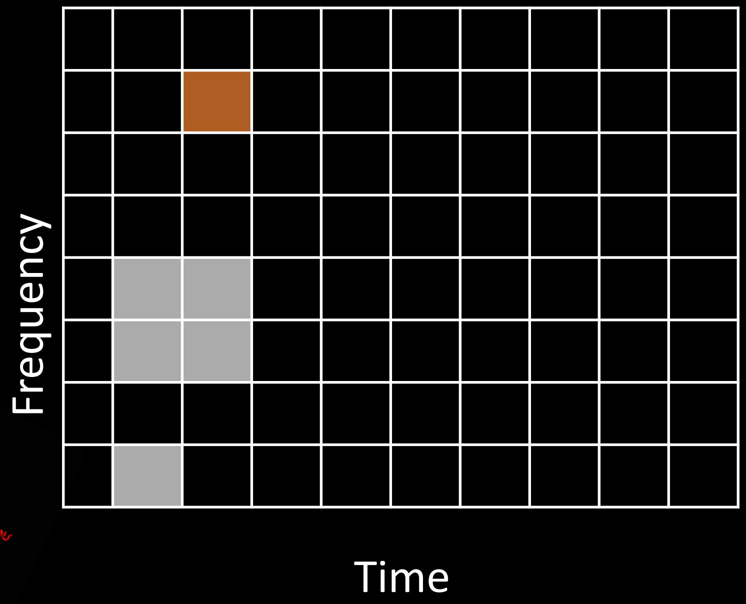
Spectrum awareness: control



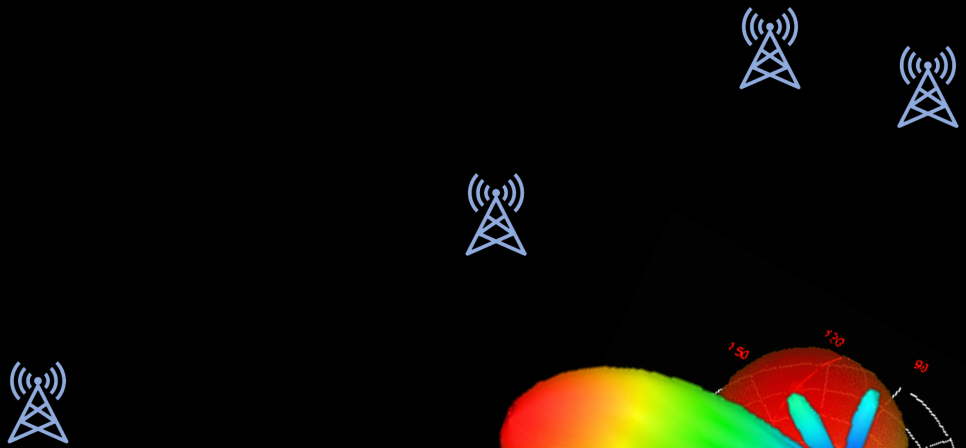


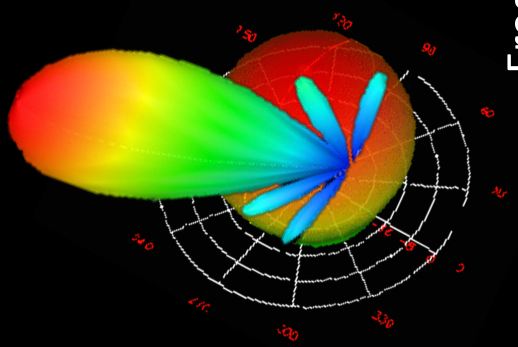
Spectrum awareness: control

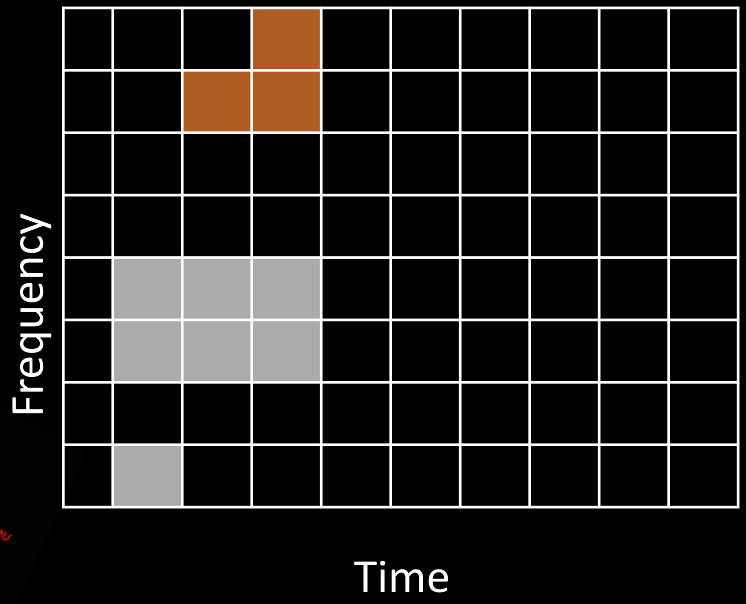




Spectrum awareness: control

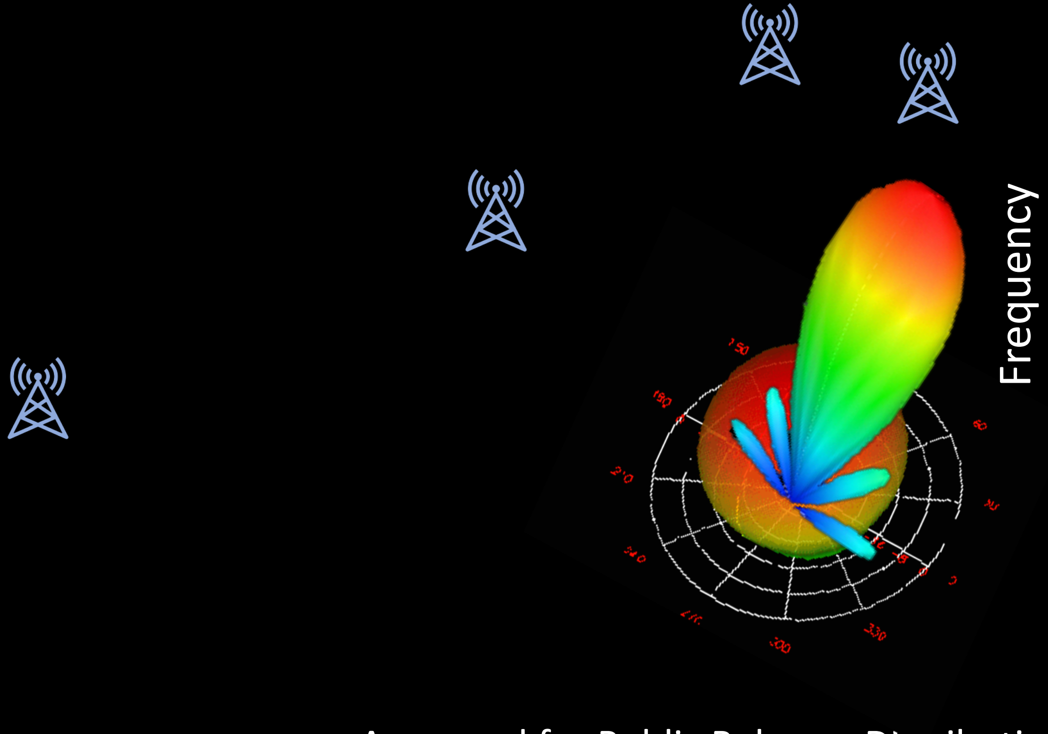


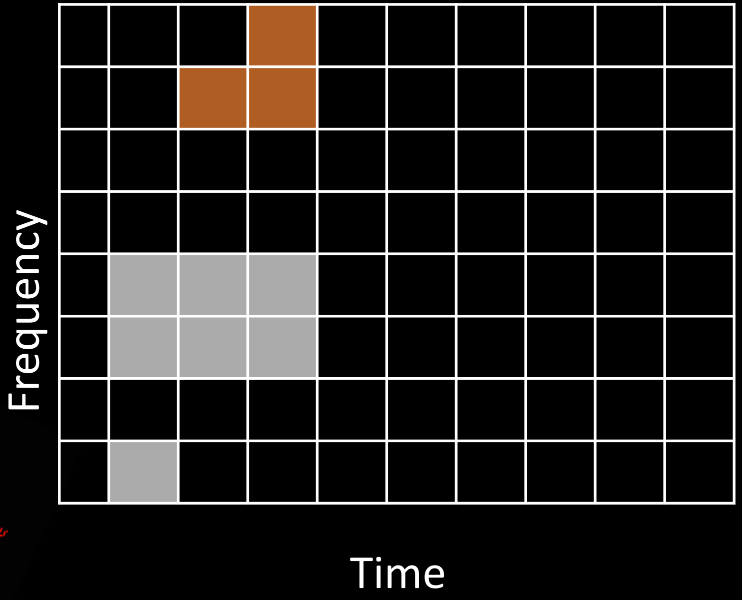




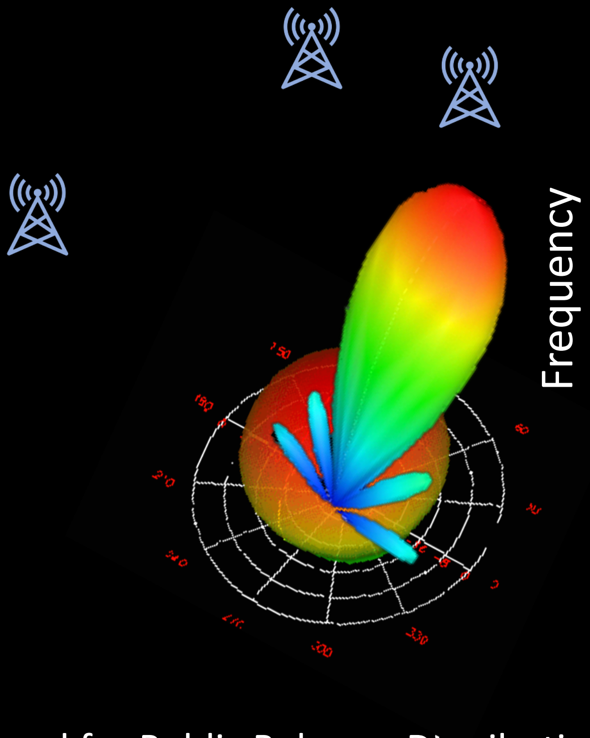
Approved for Public Release, Distribution Unlimited

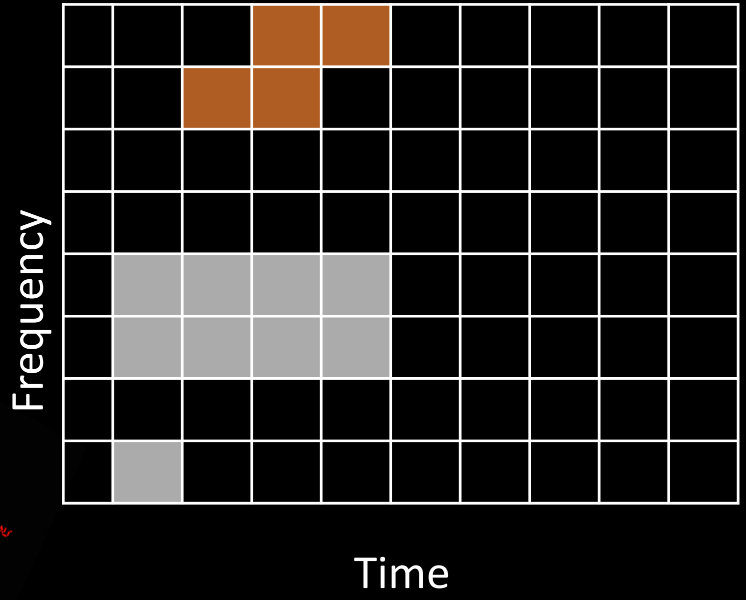
Spectrum awareness: control





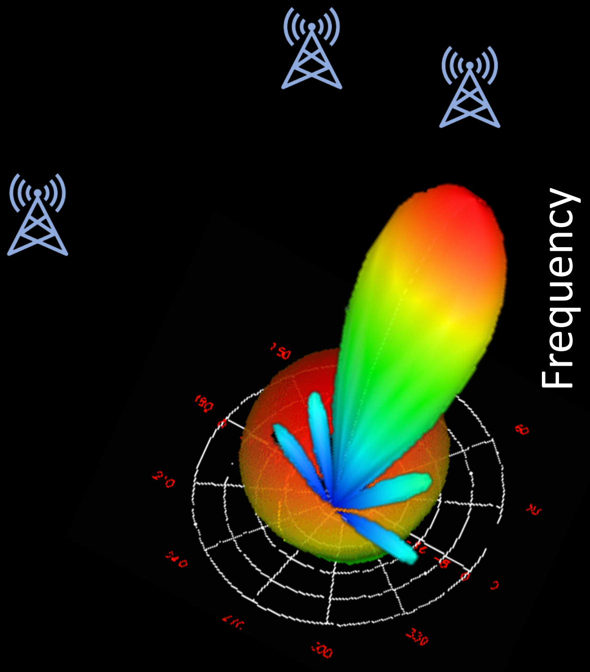
Spectrum awareness: control

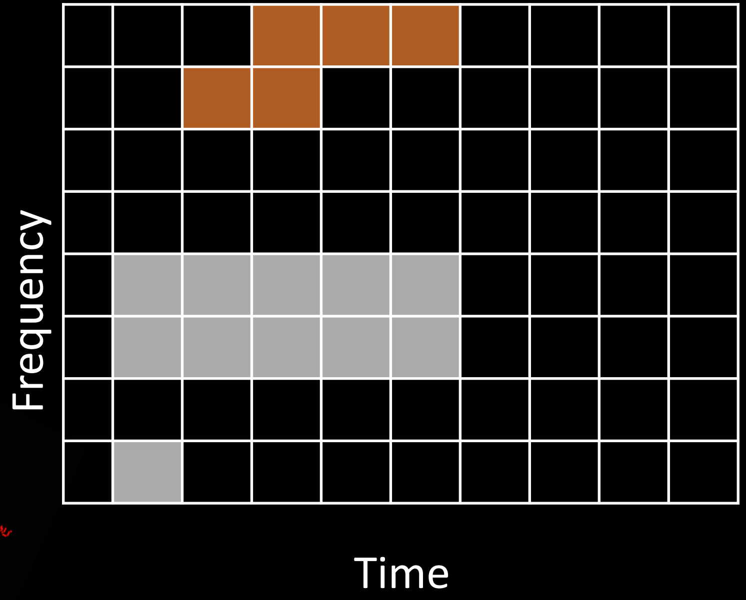




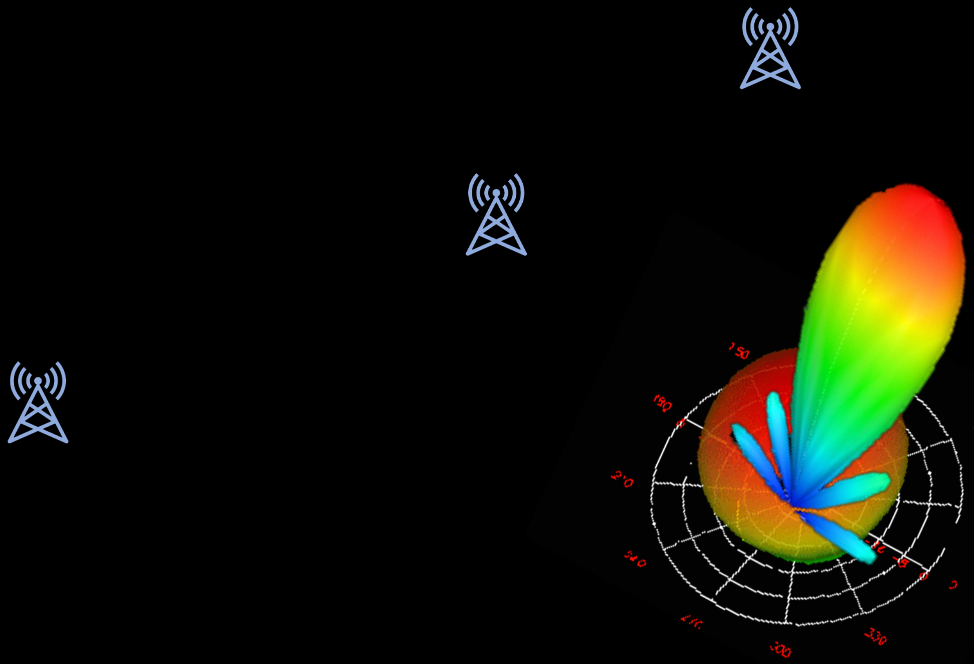
EXPEDITION TECHNOLOGY

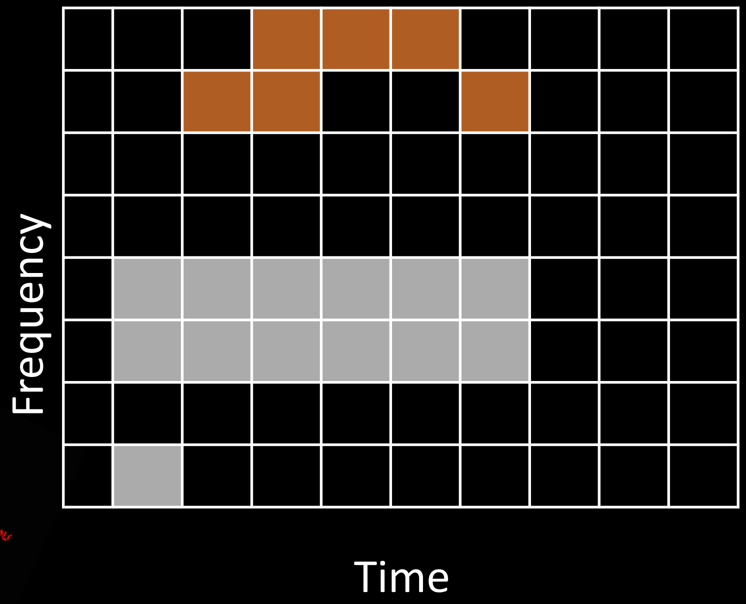
Spectrum awareness: control





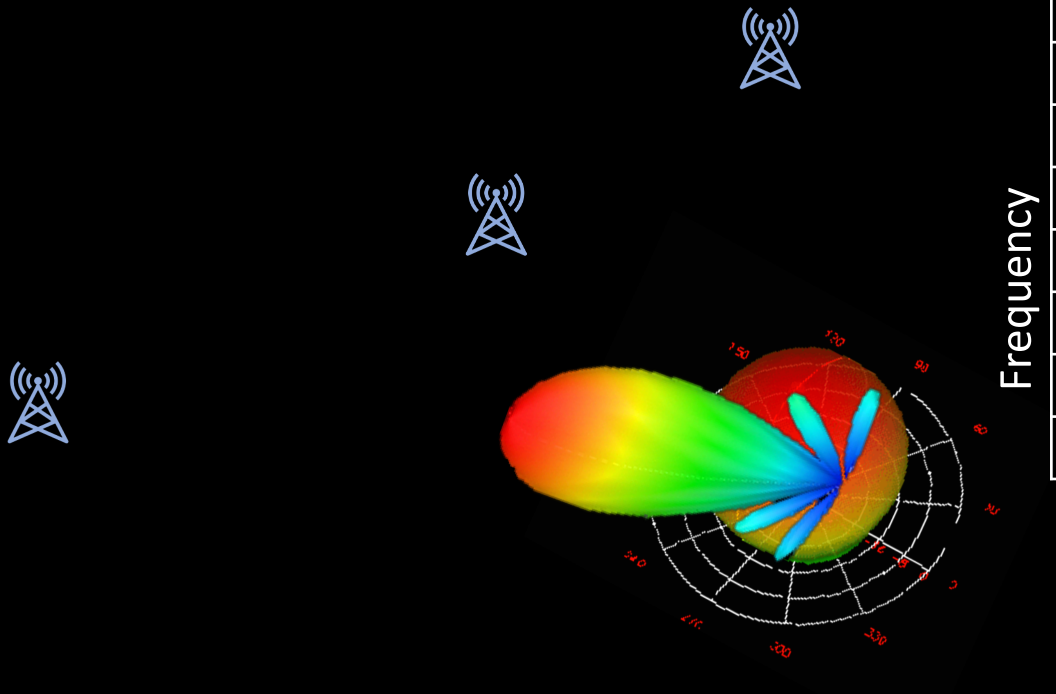
Spectrum awareness: control

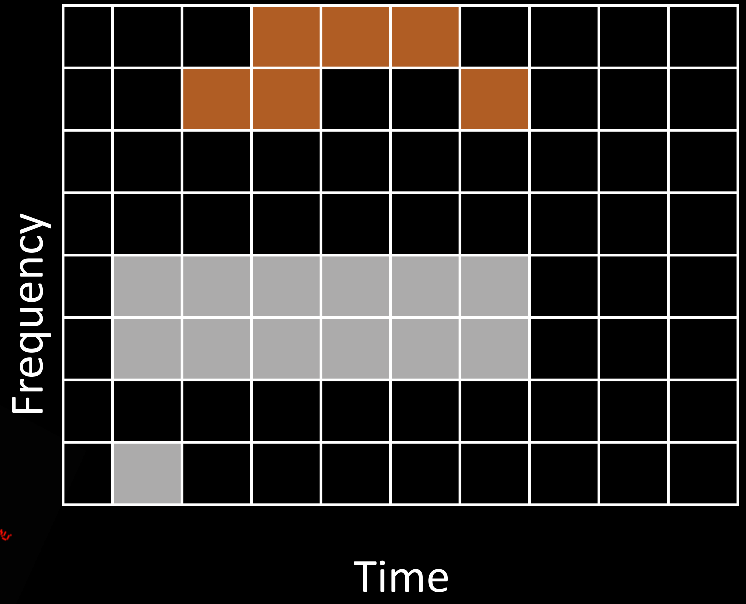




EXPEDITION TECHNOLOGY

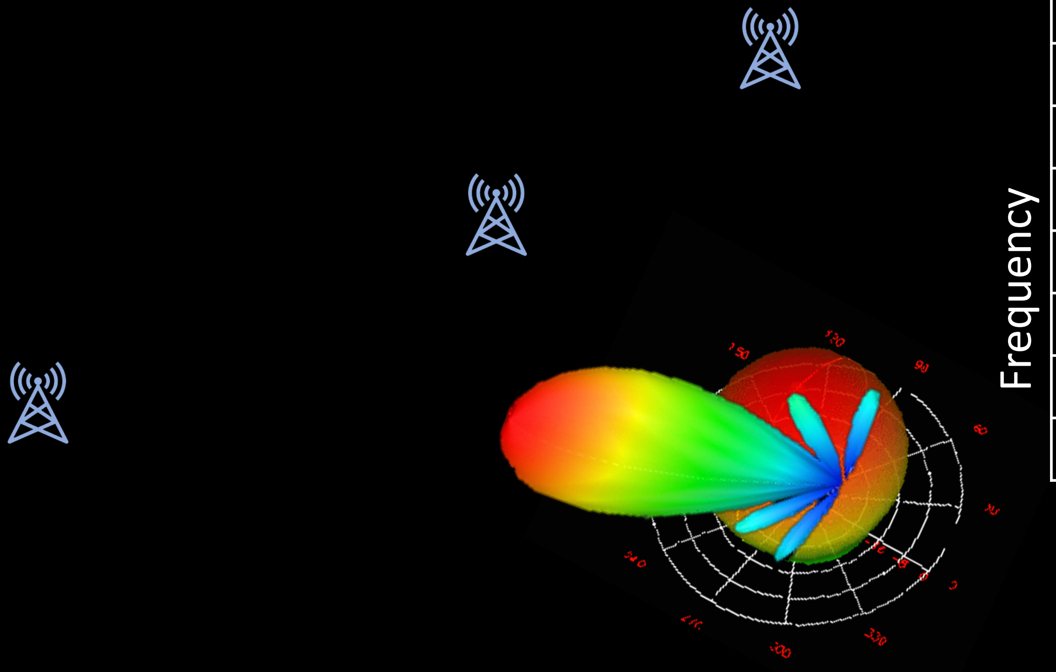
Spectrum awareness: control

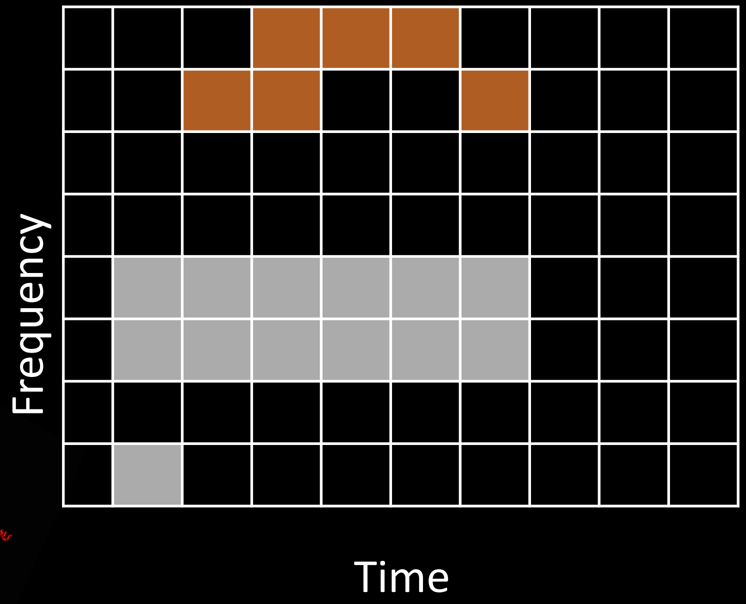




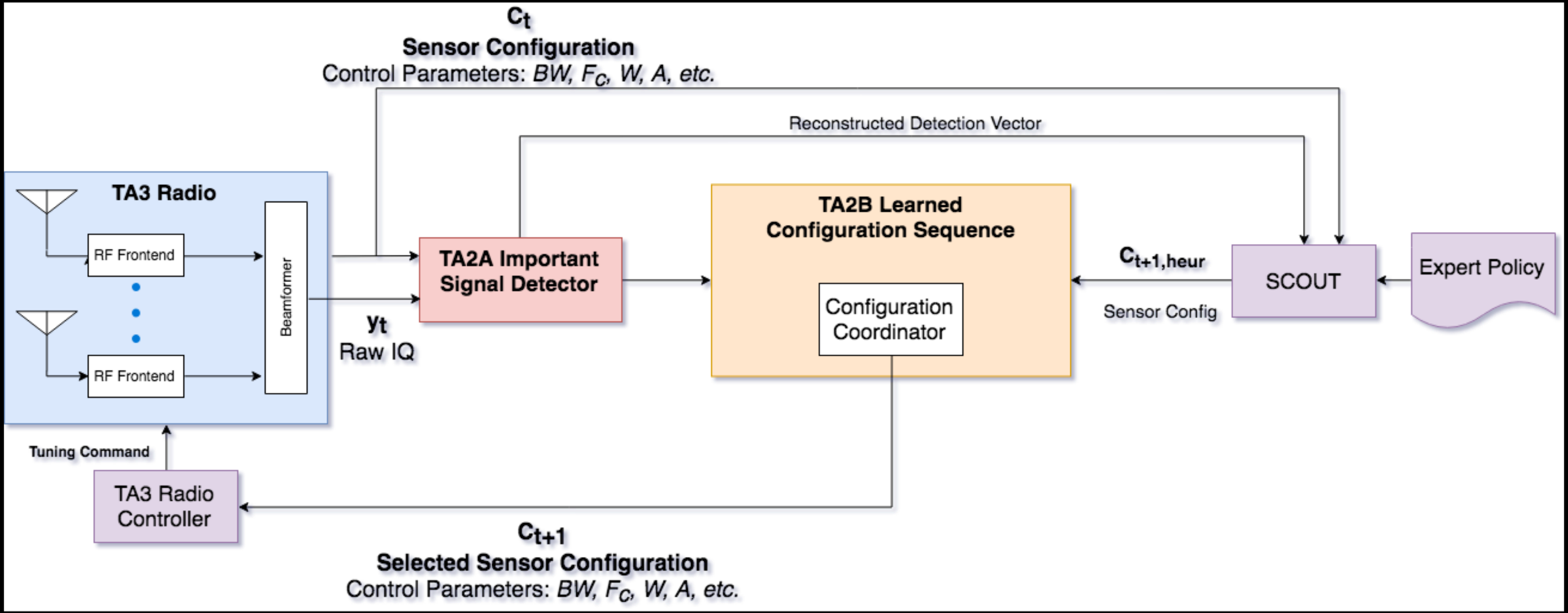
EXPEDITION TECHNOLOGY

Spectrum awareness: control





Spectrum awareness: control



Source: EXP

RFMLS datasets

https://wccftech.com/nvidia-volta-tesla-v100-ai-research/

Data rate challenges

- Sample rates 100MHz – 500MHz
- Data rates anywhere from 400MB/s to 2GB/s

https://www.nvidia.com/en-us/data-center/tesla-v100/

It would be impossible to handle such high data rates without GPUs

Signal libraries: Saliency and Control

- Dataset expected to be roughly ***60TB*** in size
- Metadata provided in SigMF format
- Backgrounds and signals of interest provided separately
- Create datasets by a “mix-and-match” scheme

Example important signals:

- 802.11x
- ADS-B
- Bluetooth
- APCO 25 Ph1
- WiMax

https://github.com/gnuradio/SigMF

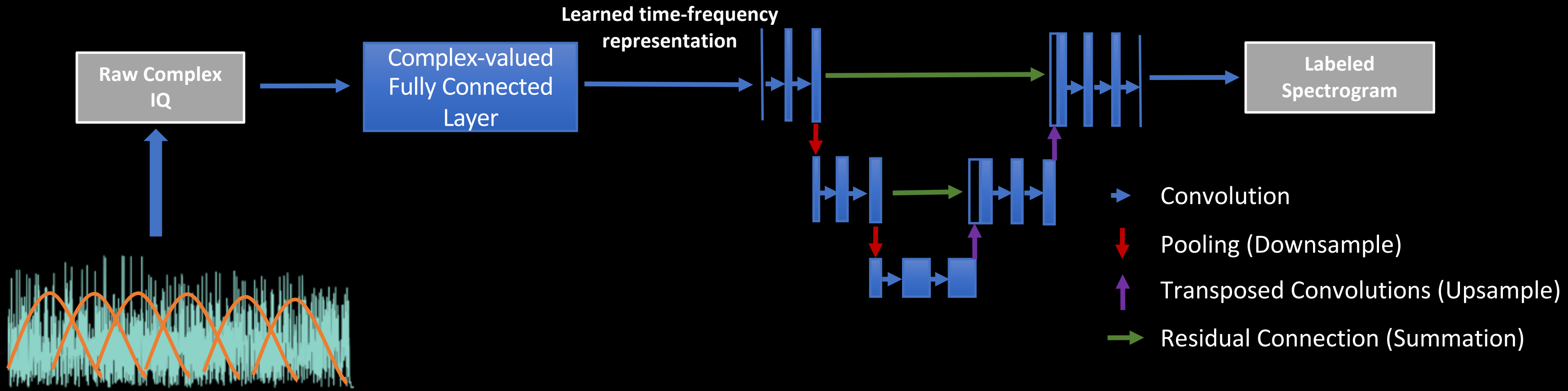
The dataset supports the following tasks:

- Identify important signals which may be possibly immersed in a complex background of confusors
- Identify signals deemed important because they are modifications of known signals
- Unsupervised anomaly detection
- Object classification from signal ensemble

Experimental results

Experimental results: saliency

Recall the network architecture:



The input is an array of dimensions 256 x 64, that is, 256 windows of 64 IQ samples each

Experimental results: saliency

Experiments performed with data simulated using GNU Radio

Experiment 1:

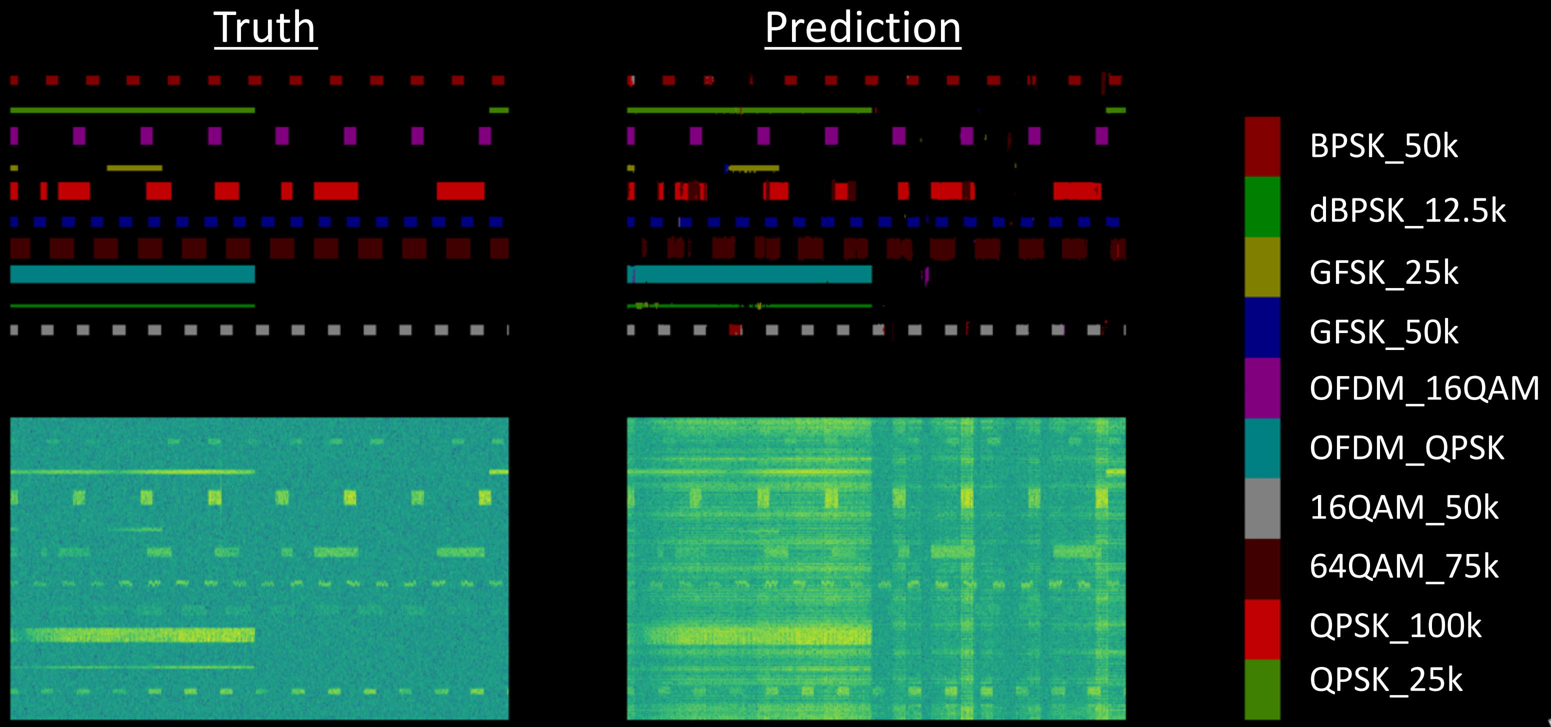
Scenario:

- 10 signal types
- 2 MSps
- “Dynamic channel”
- SNR uniformly distributed on [3dB , 20dB]
- Signal center frequency varied randomly

Training/Eval:

- Trained on 400 random configurations
- Eval on 100 unseen random configurations
- All signals labeled as “important”

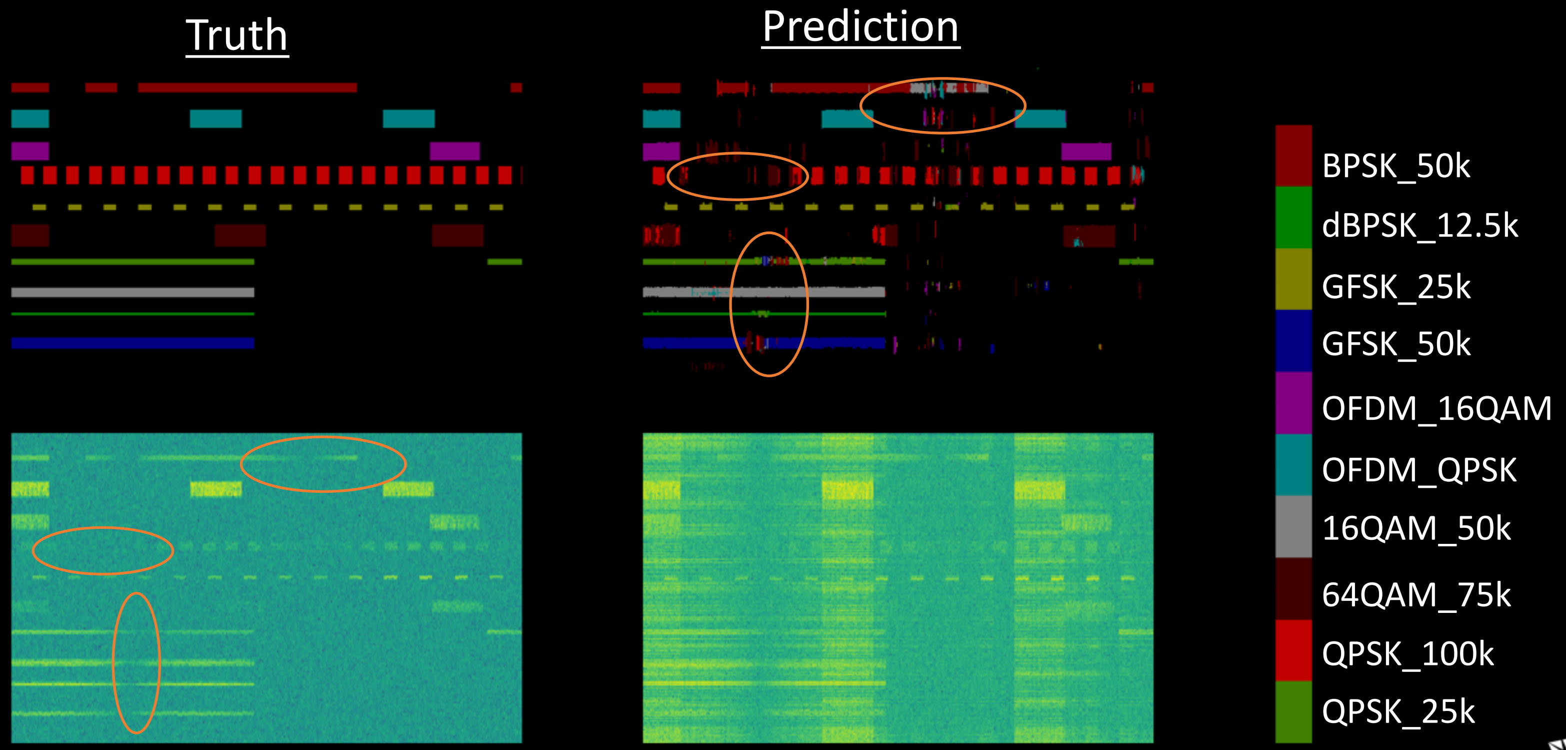
Experimental results: saliency



Source: EXP

EXPEDITION TECHNOLOGY

Experimental results: saliency



Source: EXP

Approved for Public Release, Distribution Unlimited

Experimental results: saliency

Experiments performed with data simulated using GNU Radio

Experiment 2:

Scenario:

- 10 signal types
- 2 MSps
- “Dynamic channel”
- SNR uniformly distributed on [3dB , 20dB]
- Signal center frequency varied randomly

Training/Eval:

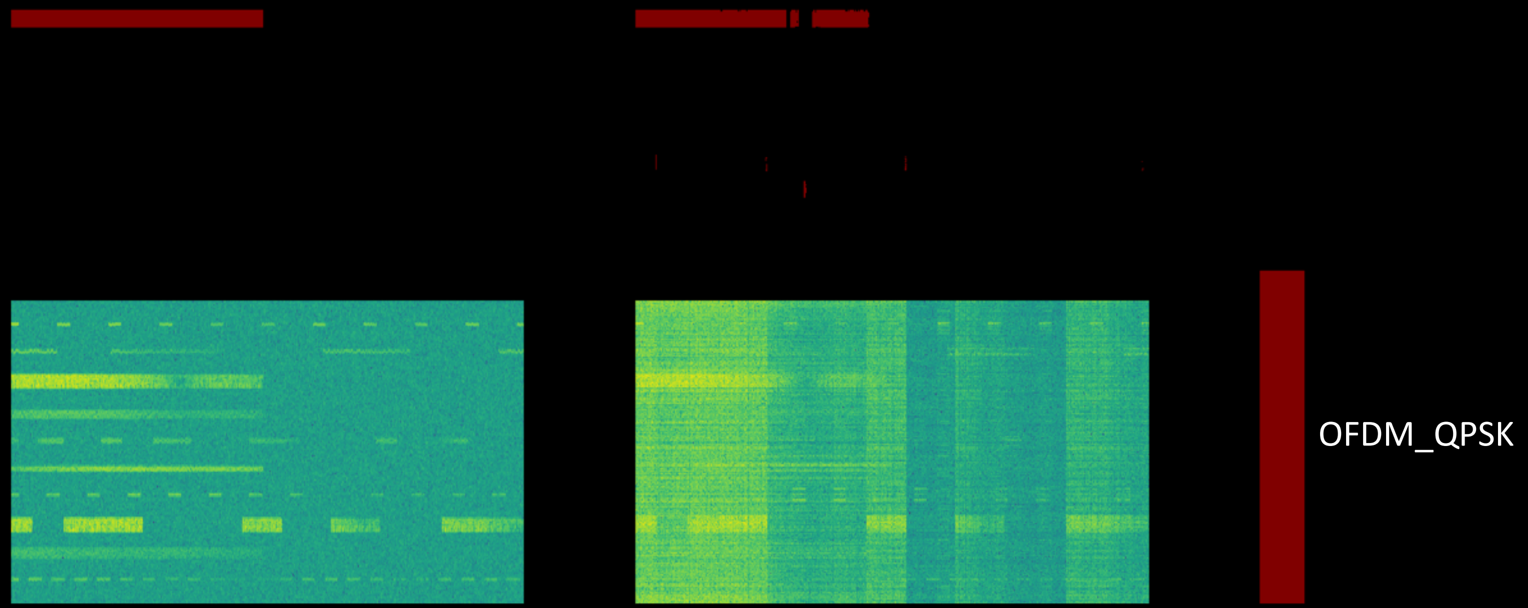
- Trained on 400 random configurations
- Eval on 100 unseen random configurations
- **Only one signal labeled as “important”**

EXPEDITION TECHNOLOGY

Experimental results: saliency

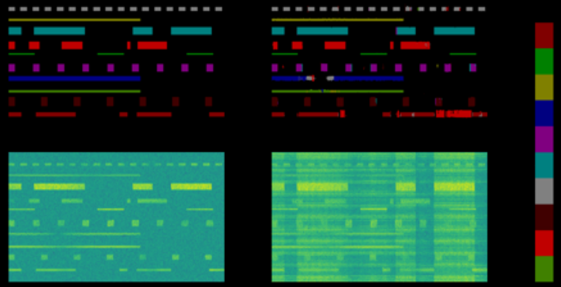
Truth

Prediction

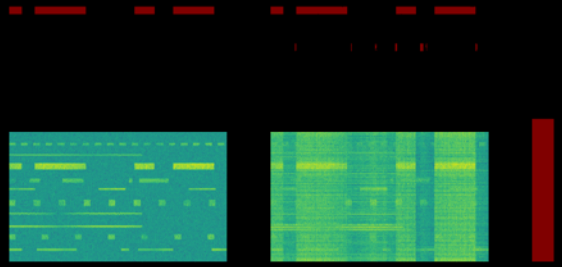


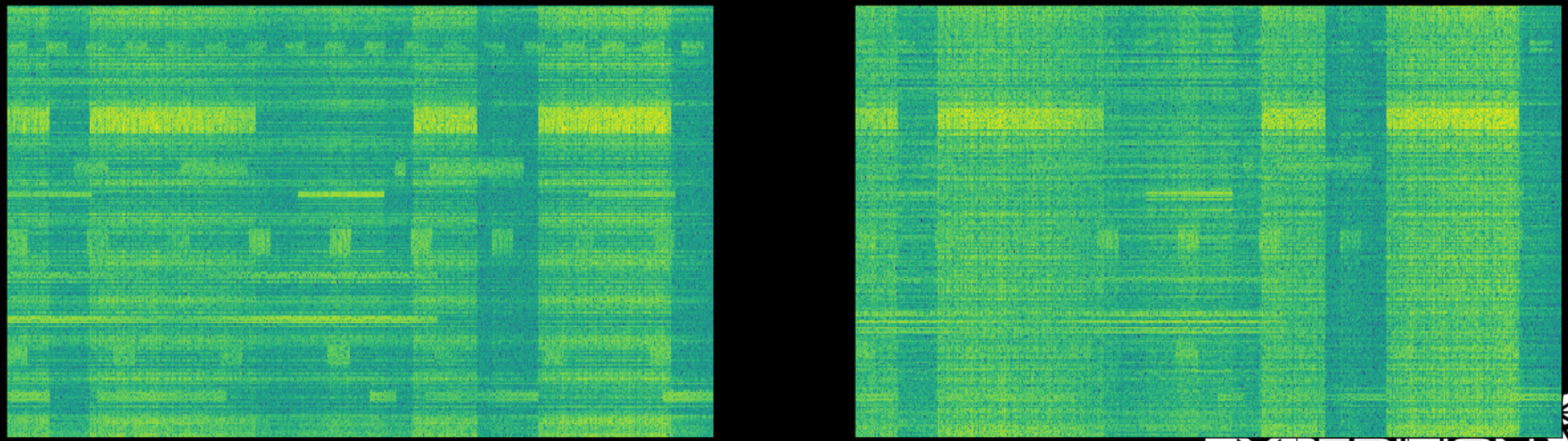
Source: EXP

Experimental results: saliency



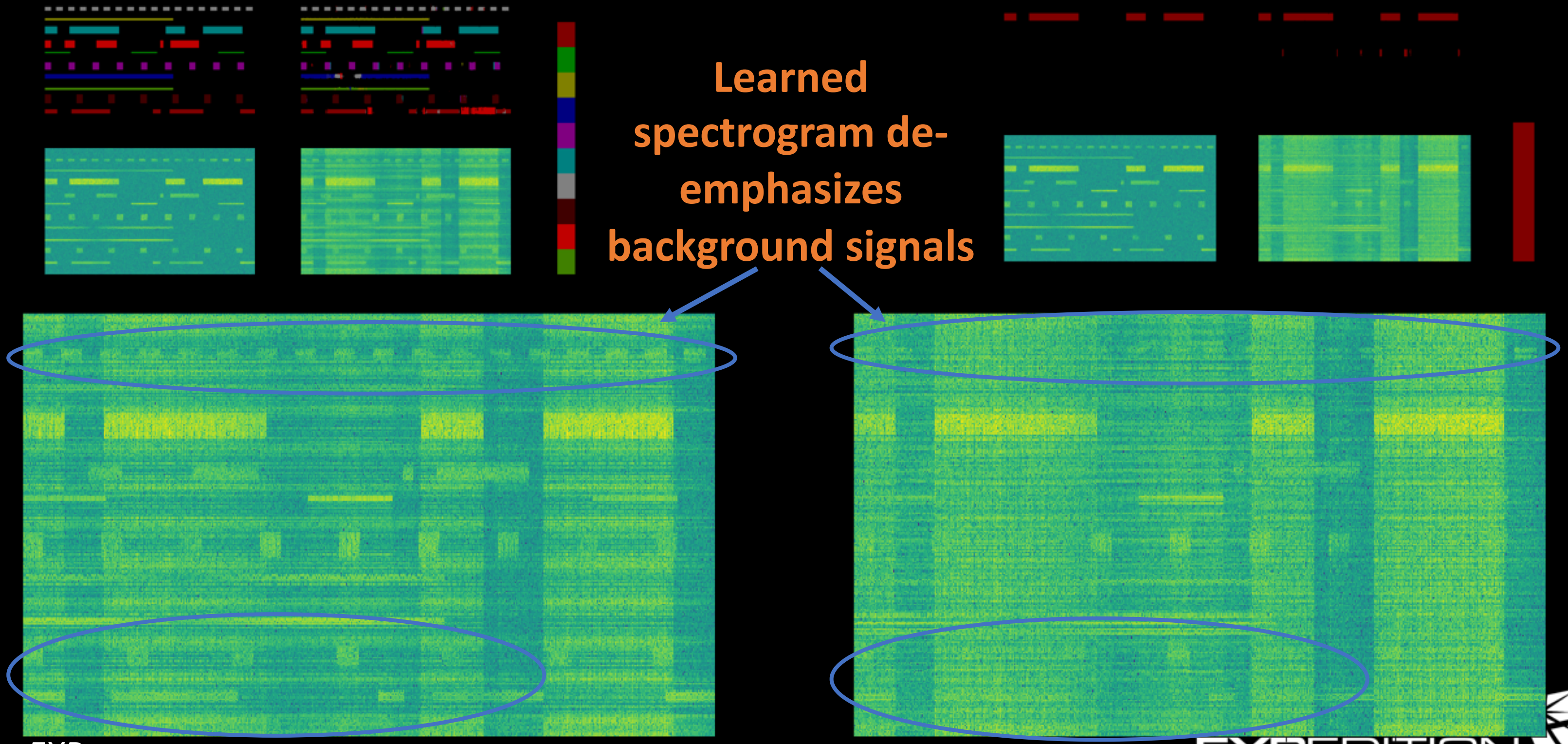
Inference on
scenarios 1 & 2
using the same data





Source: EXP

Experimental results: saliency



Source: EXP

Final thoughts

- RFMLS seeks to develop the foundational technologies to make fielded radio frequency machine learning systems a reality
- Potential to revolutionize the well-established field of wireless communications with simpler, more secure systems that can better adapt to a dynamic, increasingly challenging environment

EXPEDITION TECHNOLOGY

Contact us! (We're hiring!)

Enrico Mattei

Greg Harrison

Expedition Technology, Inc.

info@exptechinc.com

exptechinc.com

(571) 429-6141