Combinatorial-Randomness-Based Power Amplifier Datasets with RF Fingerprint Classification

Jiachen Xu, Yuyi Shen, Jinho Yi, Ethan Chen, Vanessa Chen
Carnegie Mellon University, Pittsburgh, PA 15213

IoT Security with RF Fingerprints (RFF)

The growth of the Internet of Things (IoT) brings more cyberattacks due to the large number of entry points in the network. Radio Frequency Fingerprinting utilizes features in the signals and waveforms from transmitters' unique **physical-layer** imperfections and manufacture variations to classify and authenticate devices.

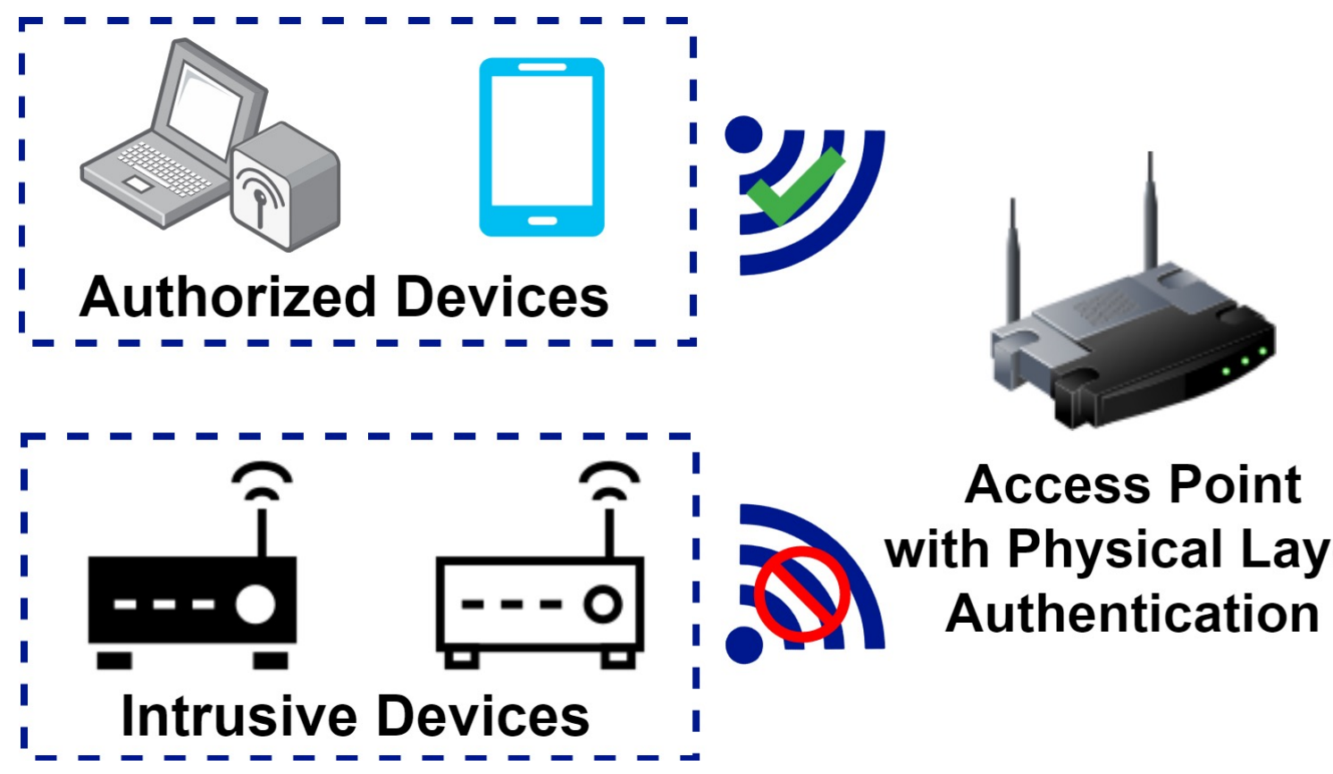


Figure 1: Wireless device authentication using physical layer RFFs.

Multiple RFFs on a Single Device

Although RFF has shown effectiveness of adding another layer of protection to the secure wireless communications, impersonators with adversarial attacks can fake the RFFs and fool this physical-layer authentication.

Power amplifier (PA) nonlinearity is one of the major factors that contribute to the RFFs on radio devices. To prevent attacks from the impersonators, combinatorial randomness (Figure. 2) [1] is exploited to augment the timestamped RFFs with a high-efficiency PA for IoT applications. By enabling different subsets of thinly sliced PA elements, the transmitter can be reconfigured with 220 subsets that exhibit distinctive RFFs to achieve a **single-device multi-RFFs** scheme. The time-stamped RFFs can effectively prevent attacks from impersonators.

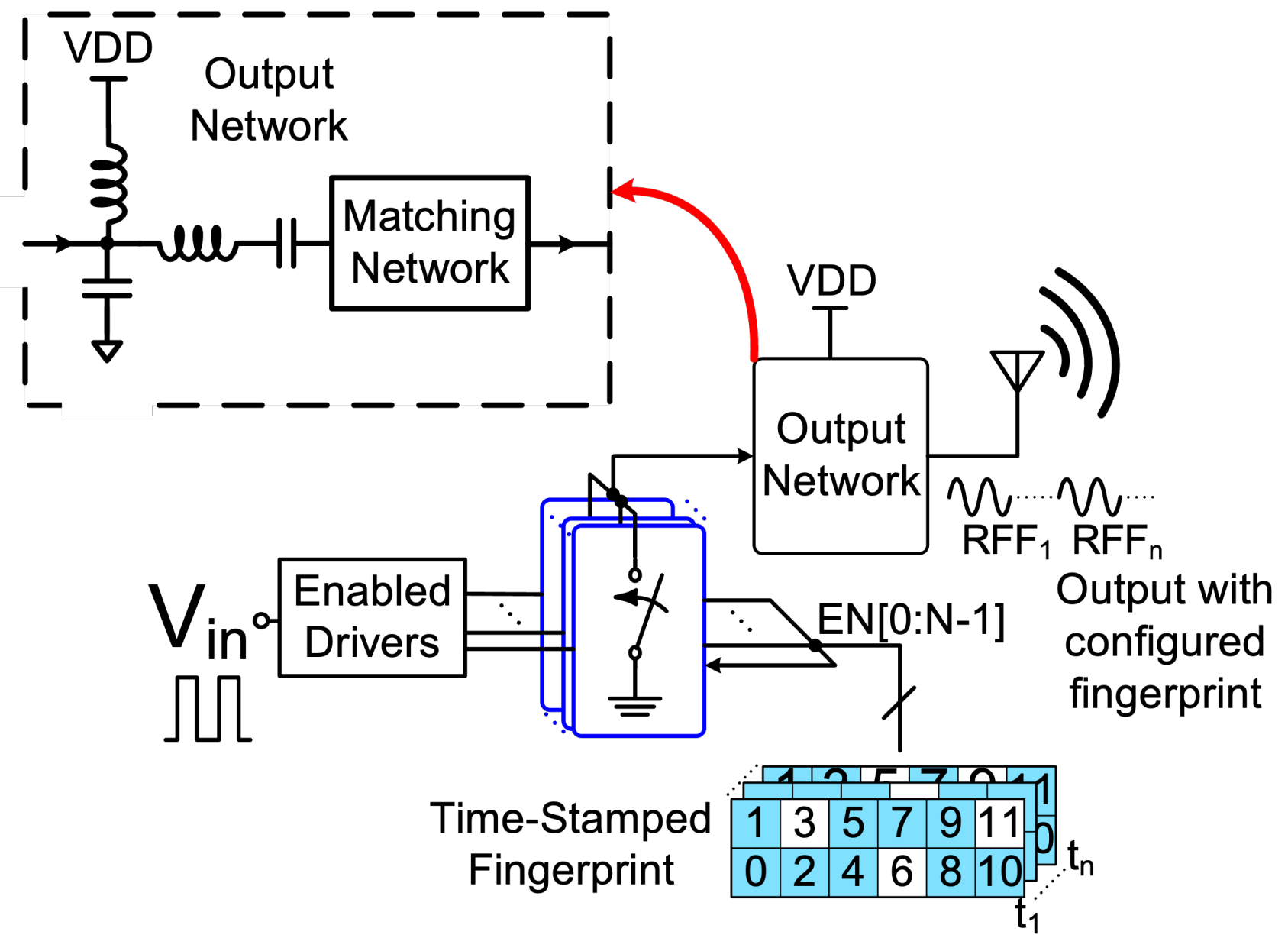
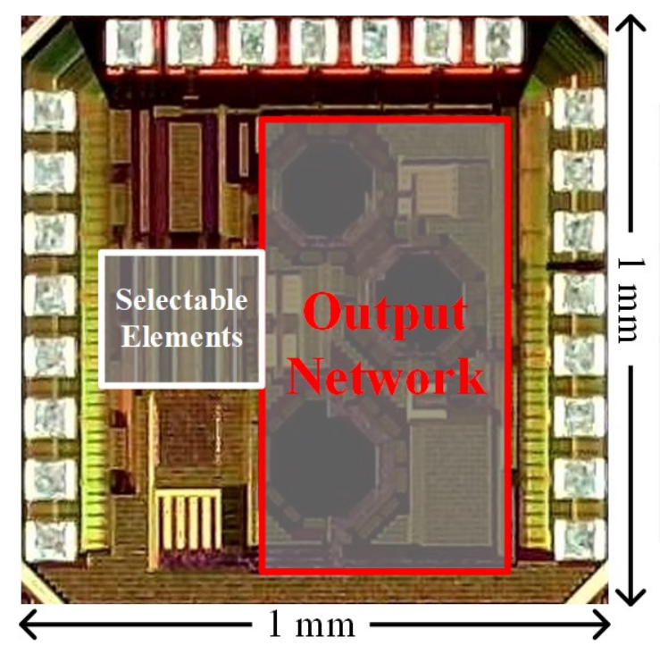


Figure 2: Combinatorial-randomness-based PA for single-device multi-RFFs.



Technology	65nm CMOS
Chip Area	1 mm^2
Core Area	0.267 mm^2
Set Number of Choosing 9 from 12	220

Figure 3: Die photograph of the combinatorial-randomness-based PA.

Dataset Collection

The baseband BLE packets are directly transmitted and received using AD9082-FMCA-EBZ. The captured I/Q samples are decimated and stored with a sampling rate of 25 MSPS to lower the required storage space. A USB-6001 NI DAQ was used to generate the control signals to switch the PA between configurations.

Over 1200 packets for each of the 220 PA configuration were collected at a baseline SNR of 35 dB. An attenuator in the measurement setup was used to lower signal power and 500 packets per configuration at SNRs of 25 dB and 15 dB were collected.

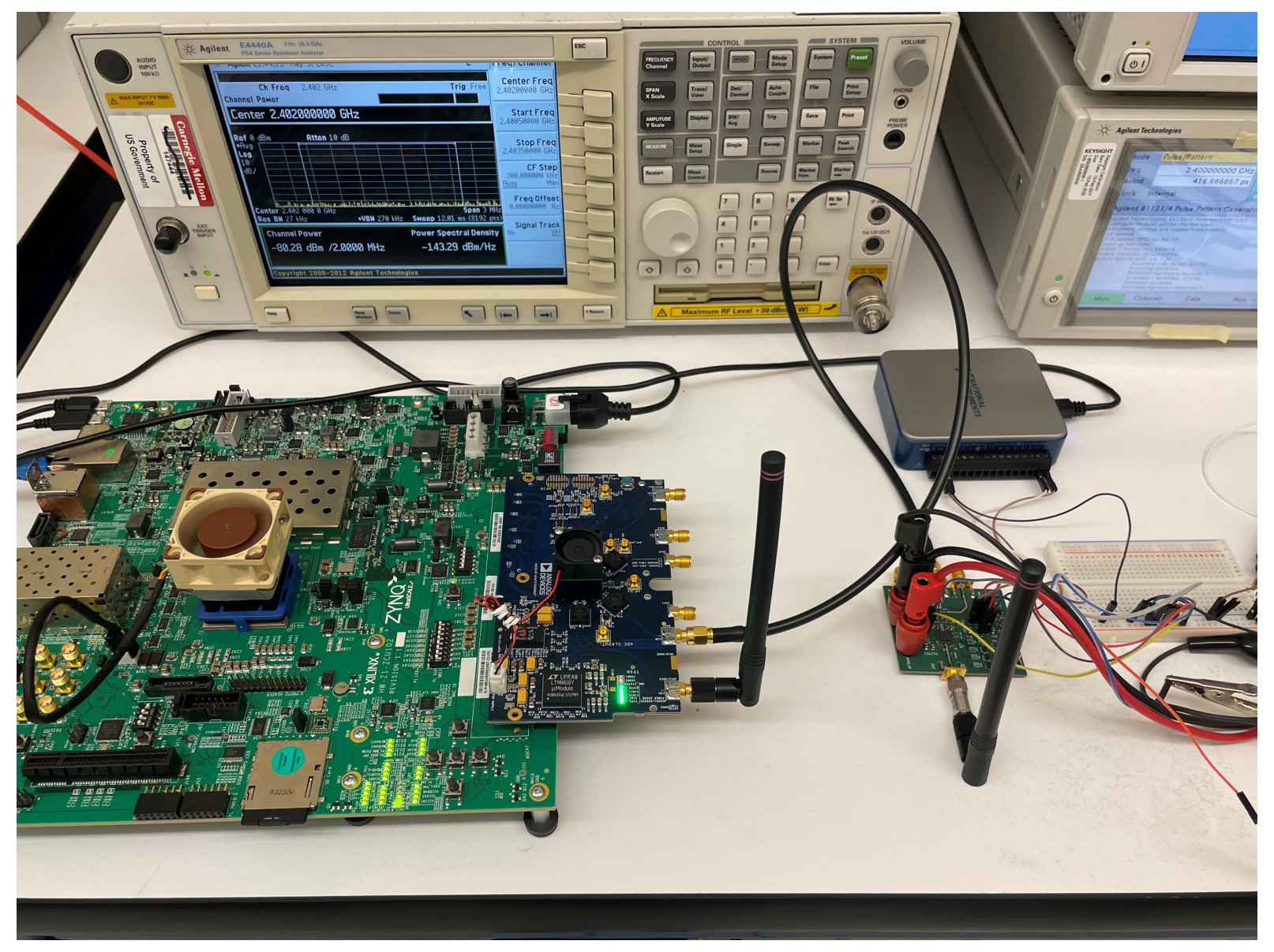


Figure 4. The equipment setup used to collect the RFF dataset. I/Q samples are transmitted using ADI AD9082 and the PA is controlled with NI DAQ USB-6001. The power spectrum of the PA output while transmitting the BLE packets was measured across PA configurations with a spectrum analyzer.

RFF Visualization

The distinct RF fingerprints produced by separate PA configurations may be visualized through the calculation of RF distinct native attribute (RF-DNA) fingerprints using the discrete Gabor Transform (DGT). Kurtosis (Kurt) and skewness (Skw) of taken from the DGT output are computed and shown in Figure 5.

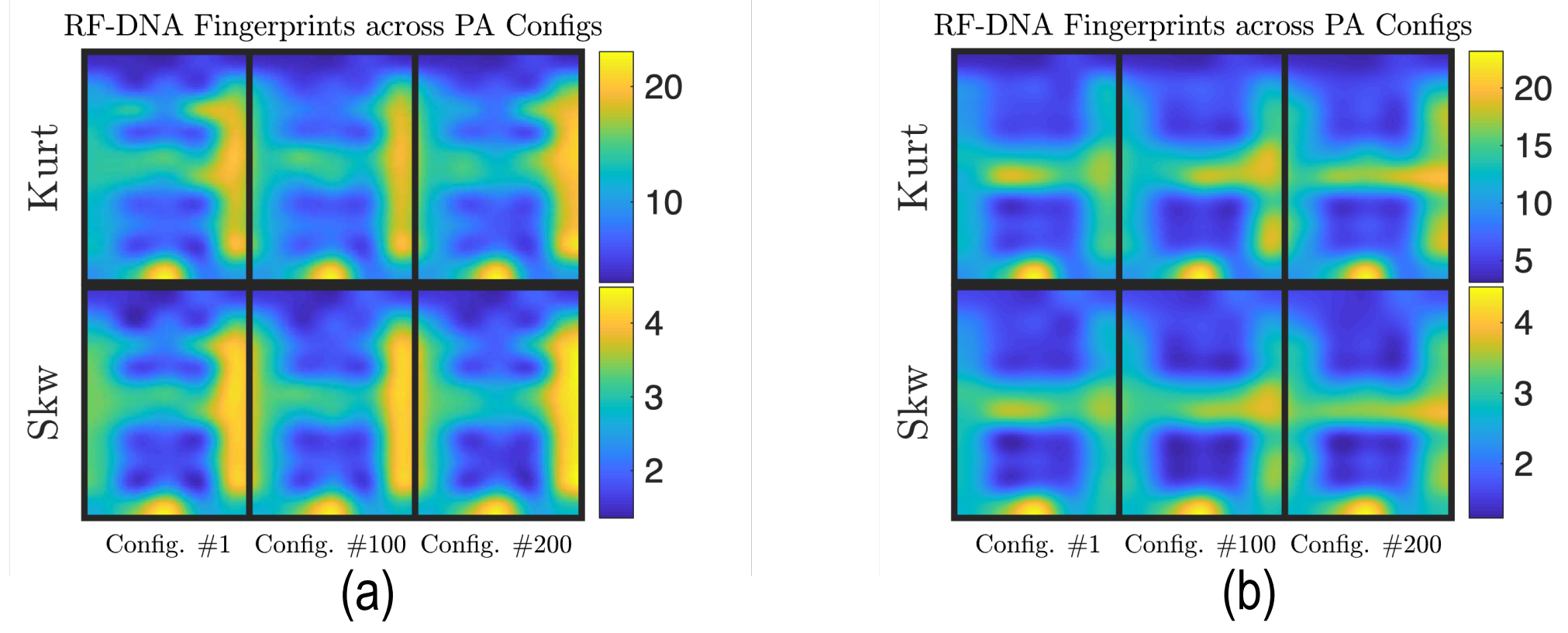


Figure 5. The average RF-DNA footprint for 3 PA configurations for (a) SNR = 35 dB and (b) SNR = 15 dB.

The full distribution of recorded RF fingerprints across PA configurations can be visualized through the usage of t-Stochastic Neighbor Embedding (t-SNE) to project the high dimensional recorded data to two dimensions for plotting Figure 6.

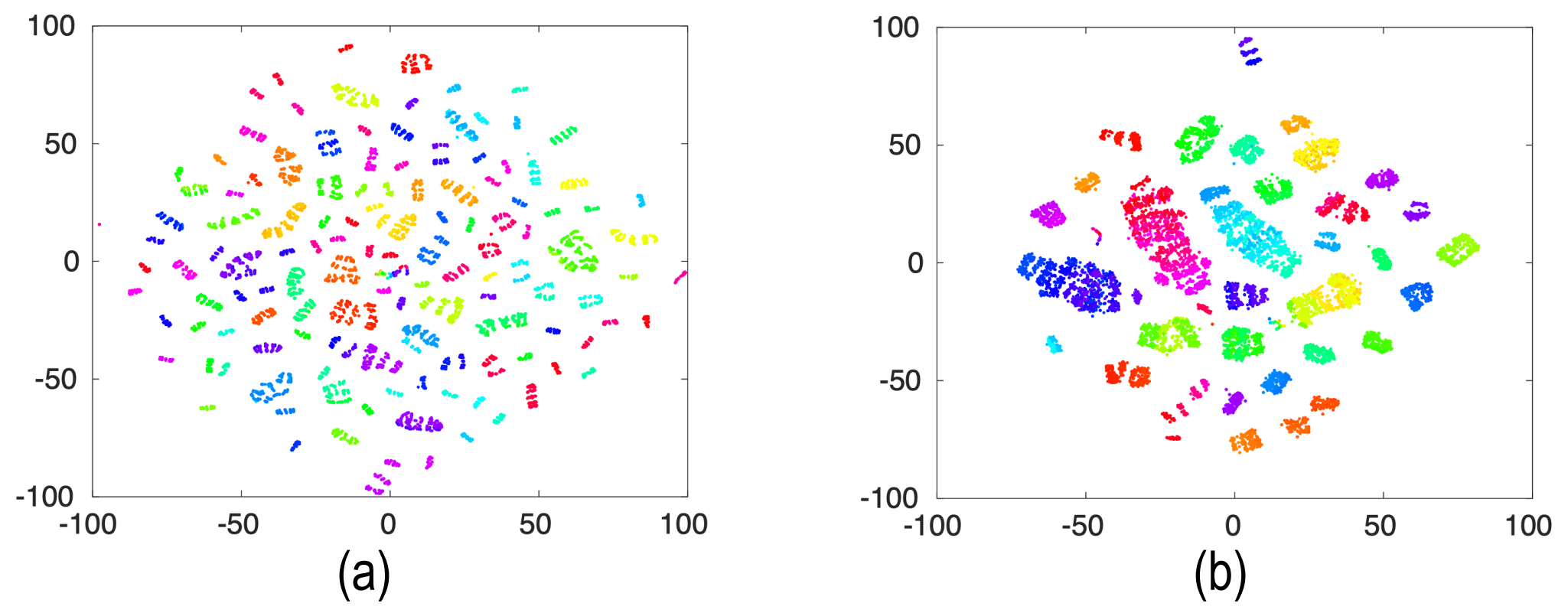


Figure 6. t-SNE was used to visualize the distribution of RFFs within the presented dataset for (a) the baseline SNR of 35 dB and (b) a moderate SNR of 15 dB.

Example Use Case of CNN For PA-Config Classification

With this dataset, an example convolutional neural network model is made to classify up to 220 PA configurations from their transmitted packets' raw I/Q data. The recorded packets are truncated to the first 40 bits of I/Q samples to isolate the fixed BLE preambles and access addresses for classification.

To estimate the receiver side's system-level hardware requirement for deploying RFF authentication, different **ADC sampling rates** and **bit resolutions** are simulated by:

- Decimating the raw data sampling rates into 1, 2, 5, 10, 15 MSPS.
- Quantizing raw data with 6, 8, 10, 12, 14, 16 bits.

The CNN is tested with each combination. The classification results are shown in Figure. 7.

The SPS=5 and bit resolution=10 model, which achieves 98.53% accuracy with 220 classes (PA-configs) is selected to carry on further analysis for its moderate hardware requirements.

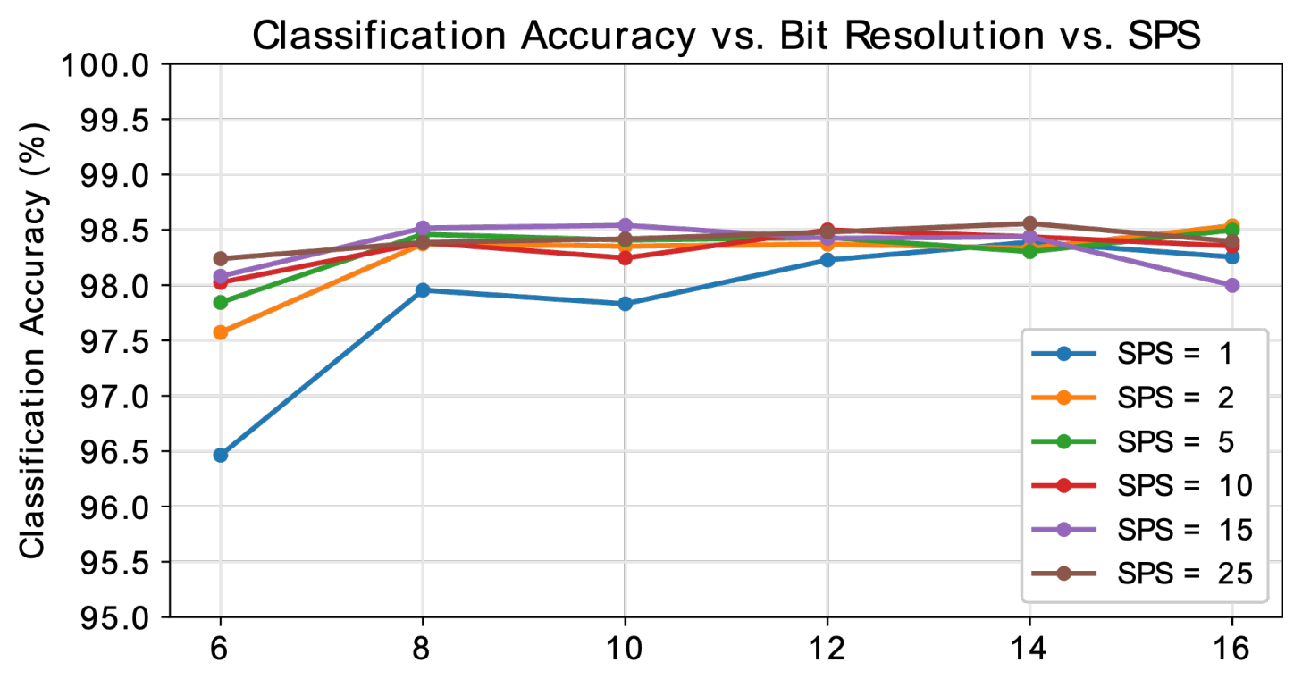


Figure 7: The CNN model's 220-configuration RFF classification accuracy with different receiver sampling rates and bit resolutions

Table 1: Example CNN Architecture

Layer	Output Dimensions
Input	2 × input size (in)
Conv, 16 Ch 1×5, stride = 2	16 × in/2
Conv, 16 Ch 1×5, stride = 2	16 × in/4
Linear	128
Linear	220

It is expected that some configurations in the PA would exhibit similar RFFs, thus affecting the classifier's ability to distinguish between these configurations (Figure 8).

Examining the confusion matrix and excluding less-distinct configuration could improve the overall classification accuracy (Figure 9).

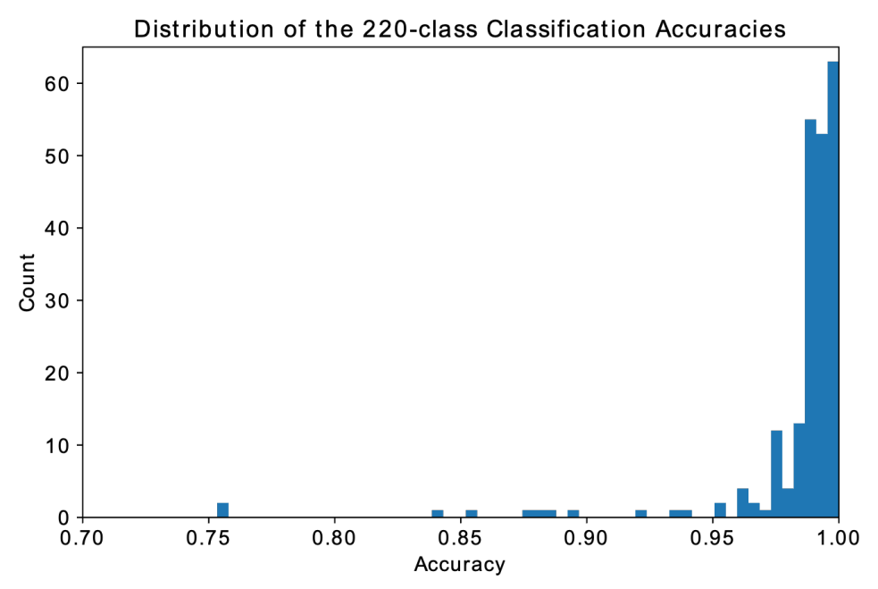


Figure 8:The distribution of 220-configuration classification accuracies.

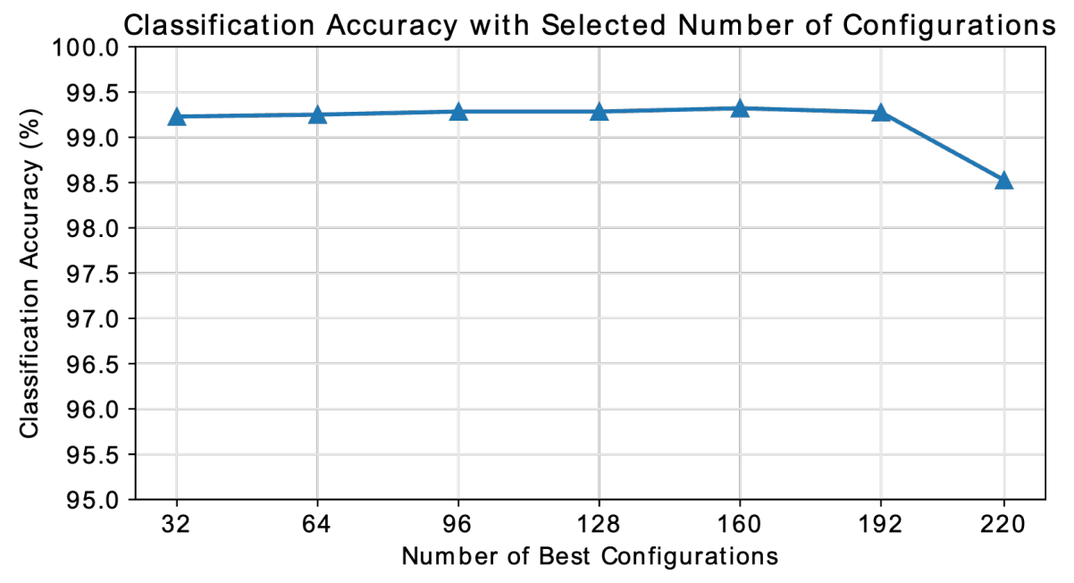


Figure 9:The distribution of 220-configuration classification accuracies.

The classifier maintains a good accuracy even when the training data is scarce (Figure. 10).

The ability of classifying PA configurations is remained when communication environment is worse (low-SNR), if noisy data is included in the training dataset (Figure. 11).

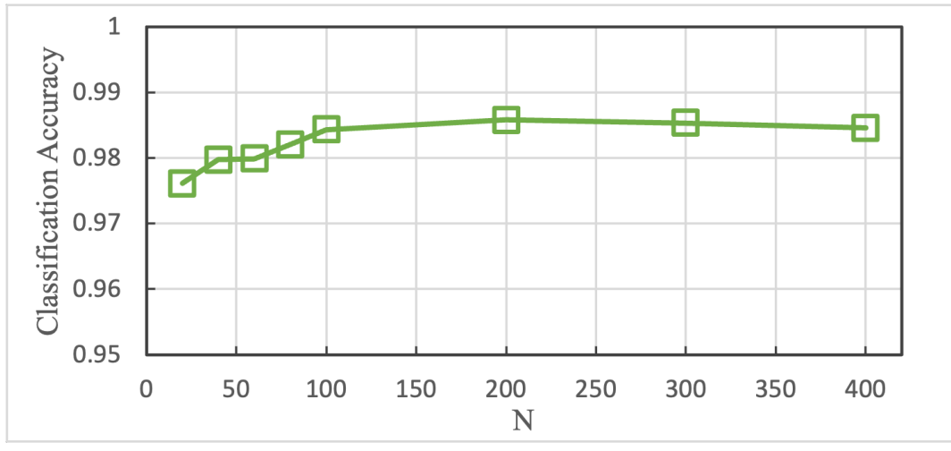


Figure 10: Classification Acc vs. Training Data size. N is the number of packets the tested training set had for each PA configuration.

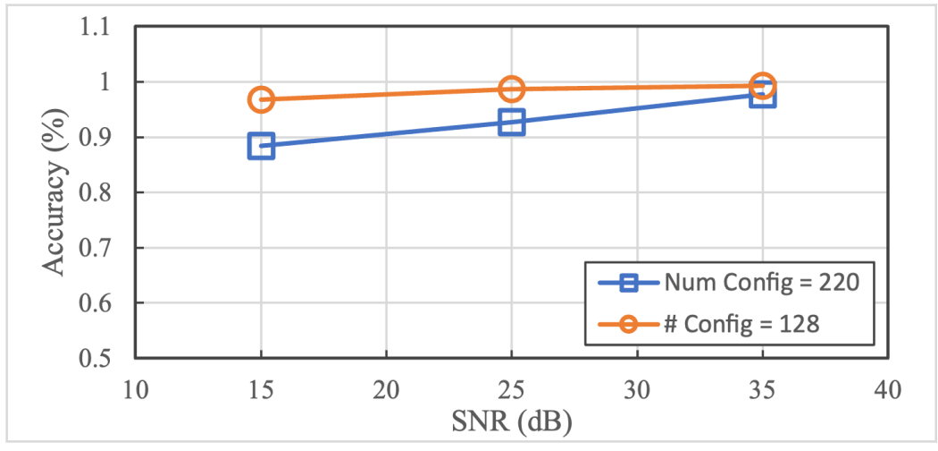


Figure 11: Classification Acc vs. SNR. The figure shows both cases for choosing 128 and 220 configurations. The tested model was trained by data with SNRs of all 15, 25, and 35dB.

CNN on Hardware

To verify the overhead of deploying RFF to a real system, the CNN model is tested on a Raspberry Pi 3B+. The trained models were also processed with post-training-quantization with TensorFlow Lite to compress the model and speed up the inference. An FPGA implementation is available in [1].

Table 2: CNN Performance on Raspberry 3B+

Model Type	Accuracy	Model Size	Inference Latency	Dynamic Power
FP16	98.53%	267KB	1.32ms	0.21W
INT8	95.09%	138KB	0.38ms	0.21W

Conclusion

In this work, we presented a dataset of RF signals recorded from a combinatorial power amplifier featuring augmented RF fingerprints. The scheme of a single device with multi-RFFs is beneficial for IoT security with RF fingerprinting through multiple configurations. The dataset includes over 1200 BLE packets for each PA configuration across all 220 PA configurations at a baseline SNR of 35 dB. Moreover, over 500 packets per configuration at SNRs of 25 dB and 15 dB were also collected. We also evaluated how a lightweight CNN model achieves different performances on this dataset with various system-level considerations including receiver cost and noise analysis, which empowers the possibility of adding another layer of protection to the wireless edge devices for secure IoT communication.

References

The detailed PA architecture and an FPGA implementation of the CNN can be found in the journal paper:

[1] Y. Shen, J. Xu, J. Yi, E. Chen and V. Chen, "Class-E Power Amplifiers Incorporating Fingerprint Augmentation With Combinatorial Security Primitives for Machine-Learning-Based Authentication in 65 nm CMOS," in IEEE Transactions on Circuits and Systems I: Regular Papers, 2022

About the Author

We are from Energy-Efficient Circuits and Systems (EECS) Lab (PI: Prof. Vanessa Chen) at Carnegie Mellon University.

Our research is focused on low-power cognitive interfaces for world-to-information computing. Our work spans the design of high-performance data converters, bio-inspired computing, ubiquitous sensory interfaces, as well as hardware-based cybersecurity.

Carnegie Mellon University

ENERGY-EFFICIENT CIRCUITS & SYSTEMS LAB

Acknowledgements

This work was supported by the National Science Foundation under Grant No. 1952907, 1953801, and 2028893.