Spectrum Monitoring: A Novel Approach using Statistical Signal Processing

1st Gaurav Duggal
M.Tech
IIIT Delhi
New Delhi, India
gaurav17091@iiitd.ac.in

2nd Kaustubh Singh
B.Tech
IIIT Delhi
New Delhi, India
kaustubh15143@iiitd.ac.in

3rd Shubham Khanna
B.Tech
IIIT Delhi
New Delhi, India
shubham15179@iiitd.ac.in

4th Shubhankar R Butta
B.Tech
IIIT Delhi
New Delhi, India
shubhankar15180@iiitd.ac.in

***Abstract*—This document talks about estimating channel occupancy using RTL-SDR, sweeping through the frequency bin to check whether the bin is occupied and modeling it to predict whether it will be occupied or free in the future.**

***Index Terms*—Cognitive radio, KL divergence, Rayleigh distribution, exponential distribution, RTL-SDR, Frequency bins.**

I. Introduction

With the increased usage of mobile communications, it has become important to study the available spectrum and analyse it to make use of this resource as efficiently as possible. In this project, we monitor the spectrum used by the police on their walkie-talkies. We then proceed to fit the inter-arrival time of spectrum activity to a suitable probability distribution, and discuss the future prospects of the findings of this project.

II. Hardware

RTL-SDR (Register-Transfer Level Software Defined Radio) - This is a device which could be plugged into our computer to scan and read radio signals in the area. It can receive frequencies varying from about 500 KHz to 1.7 GHz. We have used RTL-SDR in our project to read and scan police walkie-talkie signals.

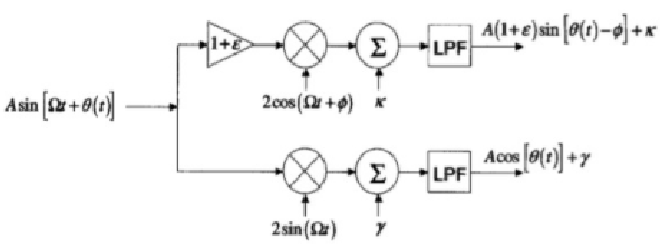

Fig. 1. Block Diagram of RTL-SDR.

Here, there is a need for error correction due to addition of forced errors like dc offset (a and b), fractional amplitude imbalance (ϵ) and phase imbalance (ϕ). These errors are then removed using IQ imbalance correction techniques used later.

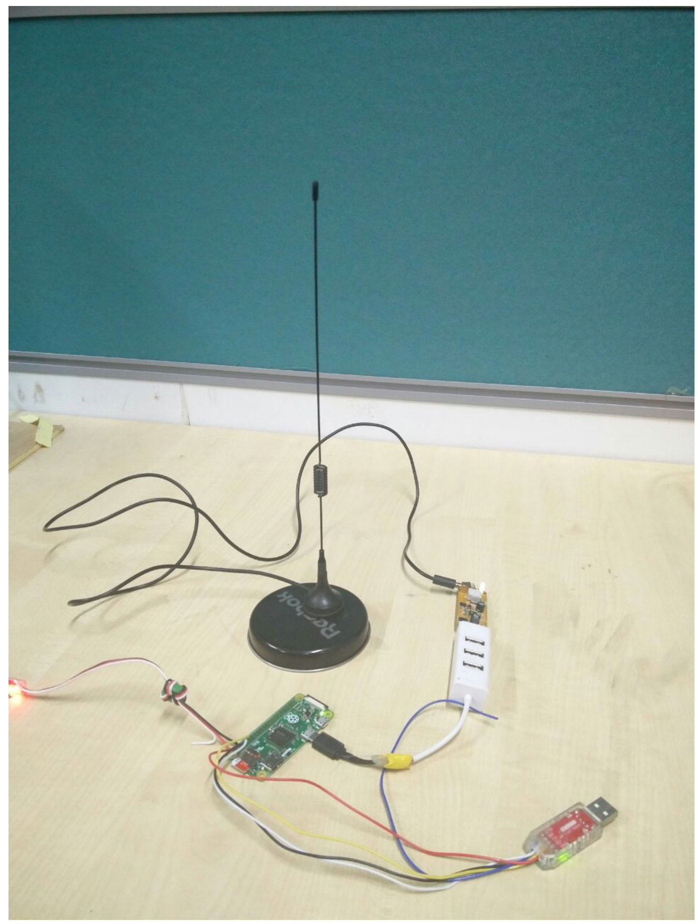

Fig. 2. Antenna Setup.

III. Data Collection

We have used RTL-SDR to collect police walkie-talkie spectral data from several places. Our data was acquired using a dongle of bandwidth 2.76 MHz. This was decimated by 15 and further sampled at 5.52 MSPS. We then applied 9 point FFT once per second and segregated it into 9 bins of 23 KHz each. The starting frequency of our data was 859.916 MHz and end point was 860.123 MHz. Thus, we had a band of 207 KHz (23 KHz * 9 = 207 KHz). We collected data from crowded

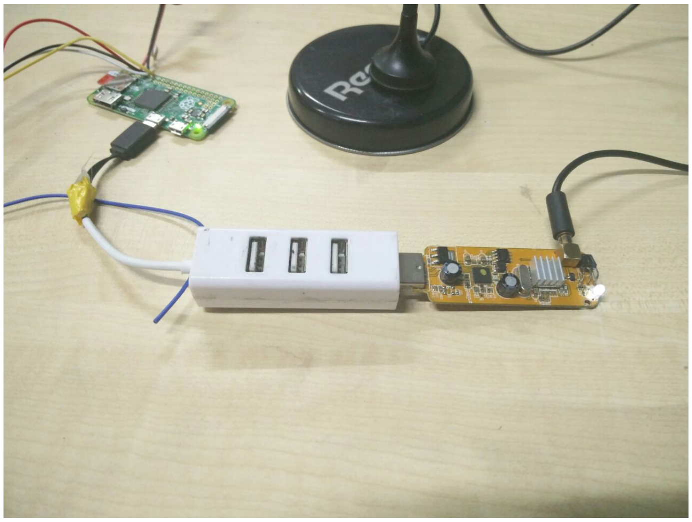

Fig. 3. Setup of RTL-SDR and RaspPi

places like Nehru Place and Govindpuri Metro station to get the best data possible.

IV. Theoretical Concepts

A. Fourier Transform

Fourier transform is a mathematical function which converts a time domain signal into its corresponding frequency domain. The frequency transform of a time domain signal gives a complex-valued function of frequency, whose absolute value gives the magnitude of that particular frequency component in the signal. Given a signal in its frequency domain, we can find out the time domain signal using inverse Fourier transform.
Fourier Transform is an extension of Fourier series for non-periodic functions. For converting a time domain signal $x(t)$ into a frequency domain signal $X(f)$ is given by

$$X(f) = \int_{\mathbb{R}} x(t) e^{-\jmath 2\pi f t}\, dt, \quad \forall f \in \mathbb{R}$$

The inverse Fourier Transform for a signal of frequency domain is:

$$x(t) = \int_{\mathbb{R}} X(f) e^{\jmath 2\pi f t}\, df, \quad \forall t \in \mathbb{R}$$

B. KL-Divergence

Kullback-Leibler divergence is a method of understanding how a given probability distribution might be different from the expected probability distribution. It makes use of information theory to calculate the divergence. It calculates a number between 0 and 1 where 0 implies very similar (if not the same) probability distribution and 1 implies the opposite or very dissimilar. It differs from the distance concept. The KL divergence of A to B is not same is the KL divergence from B to A.

$$D_{KL}(P||Q) = \int_{-\infty}^{\infty} p(x) log \frac{p(x)}{q(x)} dx$$

C. Poisson Distribution

In probability, Poisson is a discrete probability distribution. It expresses the probability of events which occur in a given time duration.
The PDF of this distribution is given by:

P(k events in interval)= $e^{-\lambda \frac{\lambda^k}{k!}}$ $program@e$

where k is the number of occurrences and λ is the expected value of the random variable.

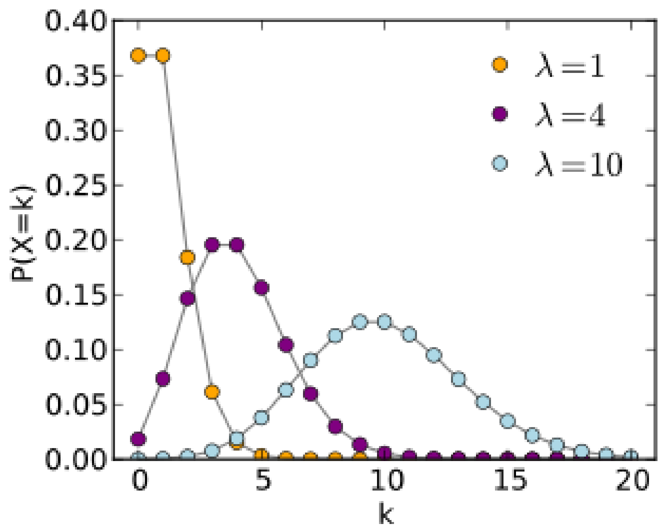

Fig. 4. PDF of Poisson Distribution.

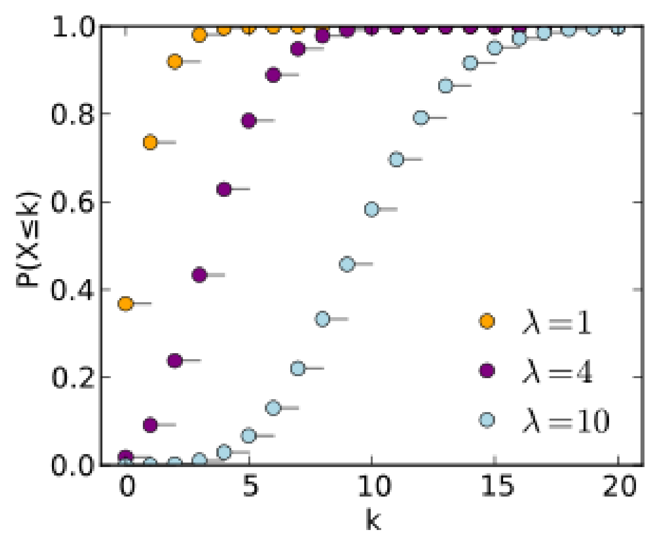

Fig. 5. CDF of Poisson Distribution.

D. Gaussian Distribution

Gaussian or normal distribution is the most commonly used continuous probability distribution. Its PMF is given by:

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where,
μ is the mean or expectation of the distribution
σ is the standard deviation, and
σ^2 is the variance.

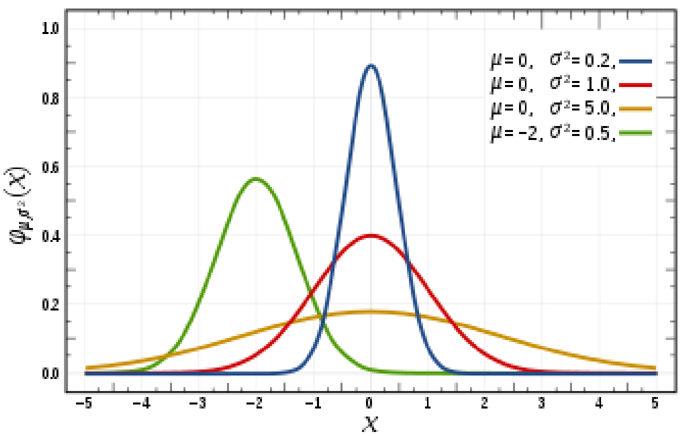

Fig. 6. PMF of Normal Distribution.

CDF of the distribution is given by:

$$\frac{1}{2}[1 + erf(\frac{x-\mu}{\sigma\sqrt{2}})]$$

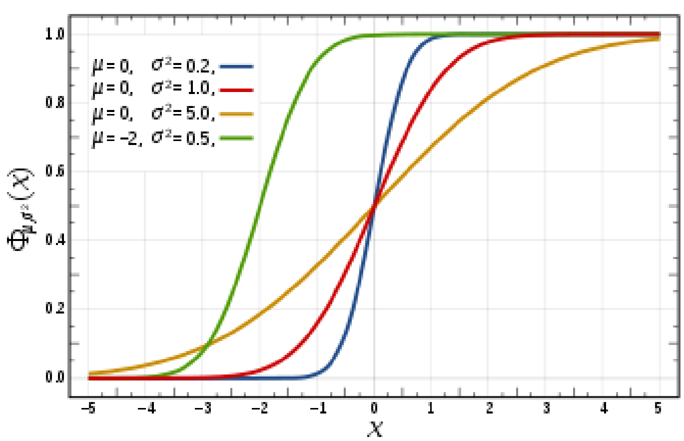

Fig. 7. CDF of Normal Distribution.

E. Rice Distribution

Rice distribution calculates the probability of a circular bivariate normal variable with a non-zero mean. The PMF of this distribution is given by:

$$f(x|\nu,\sigma) = \frac{x}{\sigma^2}exp(\frac{-(x^2+\nu^2)}{2\sigma^2})I_0(\frac{x\nu}{\sigma^2})$$

where I_0 is the Bessel function of first kind and order 0.

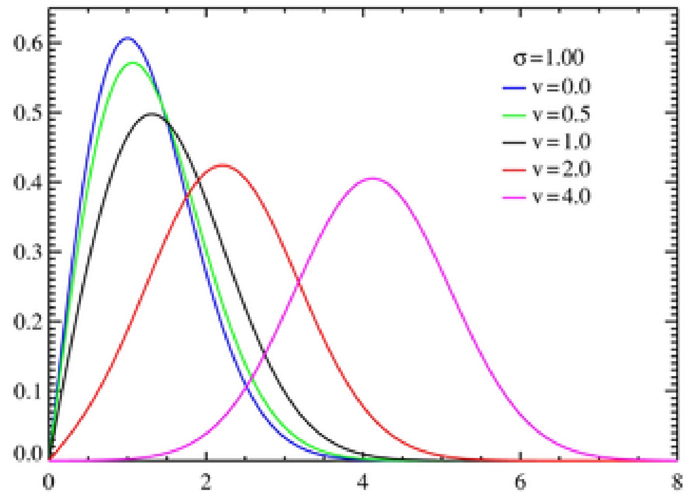

Fig. 8. PMF of Rice Distribution.

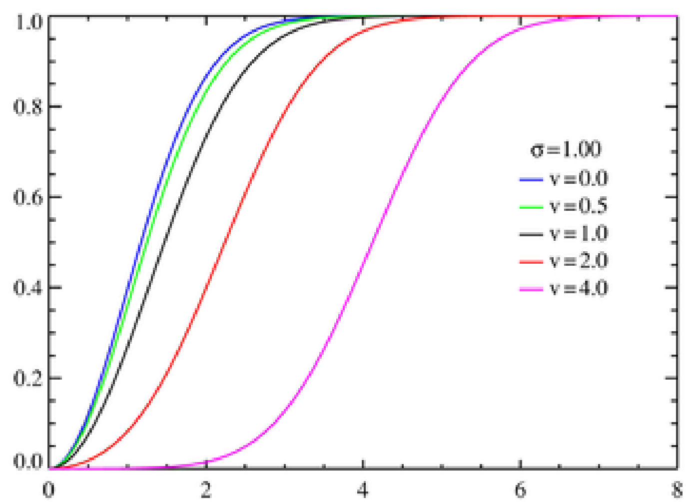

Fig. 9. CDF of Rice Distribution.

F. Rayleigh Distribution

Unlike Poisson distribution, Rayleigh distribution is continuous probability distribution for positive random variables.

$$f(x;\sigma) = \frac{x}{\sigma^2}e^{-x^2/2\sigma^2}$$

G. Exponential Distribution

Exponential distribution is the sort of distribution which defines distributions for processes that occur continuously at a constant rate (e.g. a decay process).
The PMF of exponential distribution is given by:

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

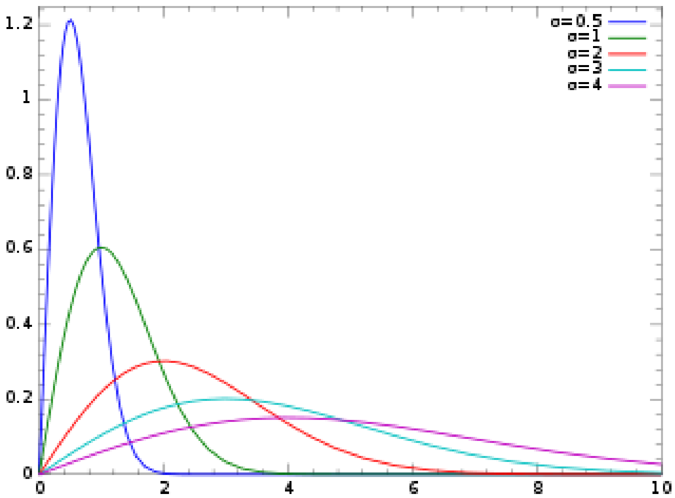

Fig. 10. PMF of Rayleigh Distribution.

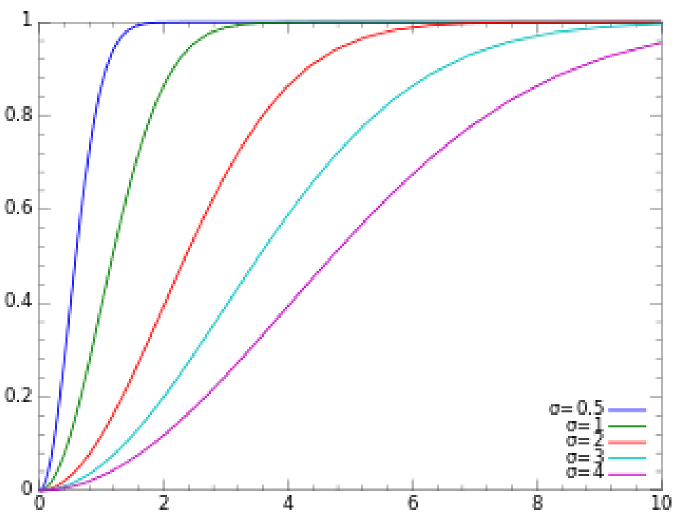

Fig. 11. CDF of Rayleigh Distribution.

where, λ is the parameter of the distribution, also known as the rate parameter.

The CDF of the distribution is given by:

$$1 - e^{-\lambda x}$$

H. IQ Imbalance

This is the method for correcting the gain and phase imbalances and the bias errors of the in-phase (I) and quadrature (Q) channels of a coherent signal processor.

With errors, the I and Q paths can be denoted by I_1 and Q_1 as given below:

$$I_1(t) = (1 + \epsilon) A \cos \omega_1 t + a$$

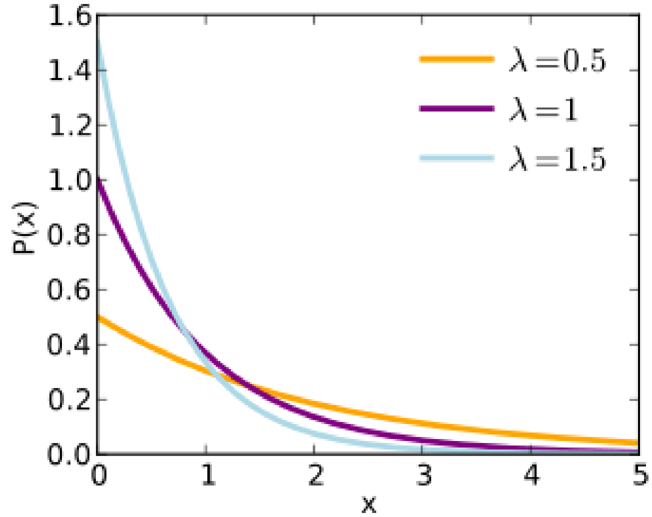

Fig. 12. PMF of Exponential Distribution.

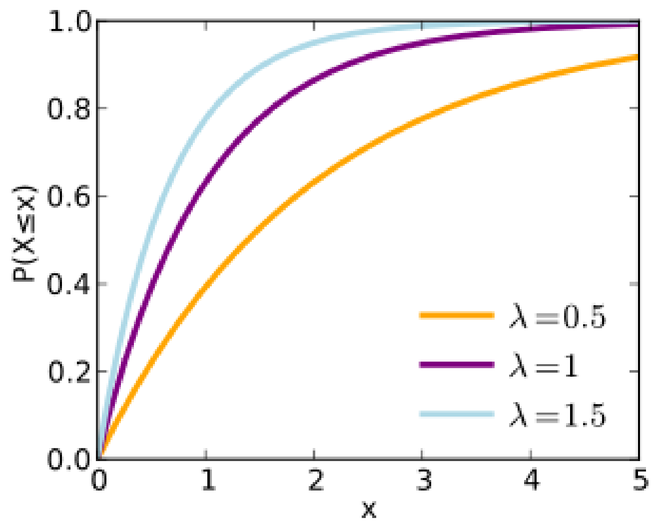

Fig. 13. CDF of Exponential Distribution.

$$Q_1(t) = A \sin(\omega_1 t + \phi) + b$$

where

ϵ is the fractional amplitude imbalance,

ϕ is the phase imbalance,

a is the dc offset in I channel, and

b is the dc offset in Q channel.

The correction of bias errors consists simply of subtracting the average level from the signal in each channel. After bias

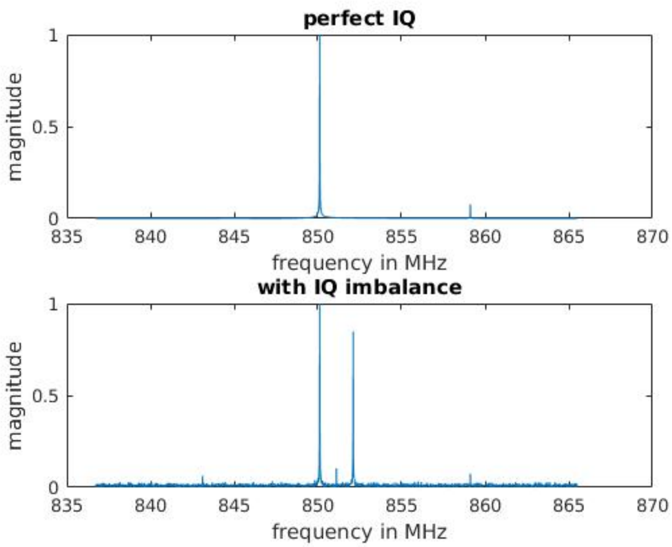

Fig. 14. Before IQ Imbalance Correction.

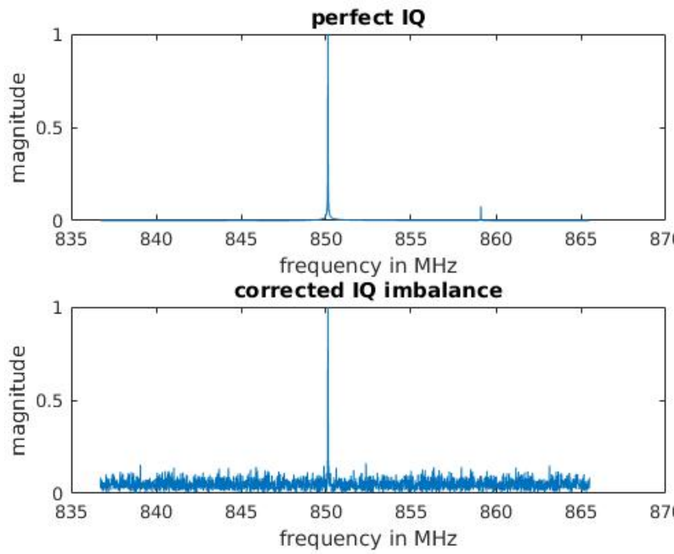

Fig. 15. After IQ Imbalance Correction.

correction the signals can be expressed as:

$$I_2(t) = (1+\epsilon)A\cos\omega_1 t$$

$$Q_2(t) = A\sin(\omega_1 t + \phi)$$

Now, there is a need to remove ϵ and ϕ from the equation. This is done using Gram-Schmidt orthogonalization procedure. The I_2 and Q_2 signals are treated as vectors and two correction coefficients P and E_1 are required to remove these errors and transform it into I_3 and Q_3 which are error free vectors. the first approach results in simpler arithmetic.
In matrix notation the signals after gain and phase correction, designated as I_3 and Q_3 , are related to I_2 and Q_2 as follows:

$$\begin{bmatrix} I_3 \\ Q_3 \end{bmatrix} = \begin{bmatrix} E_1 & 0 \\ P & 1 \end{bmatrix} \begin{bmatrix} I_2 \\ Q_2 \end{bmatrix}$$

These correction coefficients can be calculated directly in terms of time samples of test signals using filter algorithms for a 4-point DFT.

$$f(0) = x_0 + jy_0$$

$$f(T) = x_1 + jy_1$$

$$f(2T) = x_2 + jy_2$$

$$f(3T) = x_3 + jy_3$$

The filter algorithm which give the correction terms are given as follows:

$$\widehat{a} = \frac{1}{4}(x_0 + x_1 + x_2 + x_3)$$

$$\widehat{b} = \frac{1}{4}(y_0 + y_1 + y_2 + y_3)$$

$\widehat{E} = [(x_0 - x_2)(y_1 - y_3) - (x_1 - x_3)(y_0 - y_2)]/[(x_0 - x_2)^2 + (x_1 - x_3)^2] - 1$
$\widehat{P} = -[(x_0 - x_2)(y_0 - y_2) - (x_1 - x_3)(y_1 - y_3)]/[(x_0 - x_2)^2 + (x_1 - x_3)^2]$

V. Data Analysis

The data contains per-channel power given by the RTL-SDR. The spectrogram shows the calls taking place during the recording of our data. We obtain the time-stamp, sampling frequency, along with the value at a certain channel. The data is pre-processed to find a suitable threshold to check if the channel is occupied(Call is active) or not. The data analysis is now split into two segments.

1) Inter-Arrival Time
   Inter-Arrival time is defined as the time taken between the beginning of two different calls.
2) Call Duration interval
   Call Duration is the time of each call in the data we collected.

Our algorithm, is as proposed:

1) Data pre-processing: Data is processed to create each of the two observations into probability distributions and normalized. [The tail of the above distribution is cut

to have relevant data to fit in and avoid exceptional behavior of really low frequency data points at the end]
2) Choose a single simulated distribution to fit the above data.
3) Create a range of parameters of simulated distribution and, create multiple distributions corresponding to each parameter value.
4) Apply KL divergence to each simulated distribution and our distribution to find the Value of parameter which minimizes KL Divergence.

The above algorithm can gives us the best value of parameter of a single-parameter distribution, that best fits our observed distribution of data.

We ran the above algorithm on both the distributions(Inter-Arrival Time and Call duration and found the best fit Rayleigh and Exponential distributions for each.

VI. Results

Some results after Data Analysis.

TABLE I
IAT statistics

No. of Calls	1980
Occupancy Time	283 min
Percentage Occupancy	28%

A. Inter Arrival Time with Rayleigh

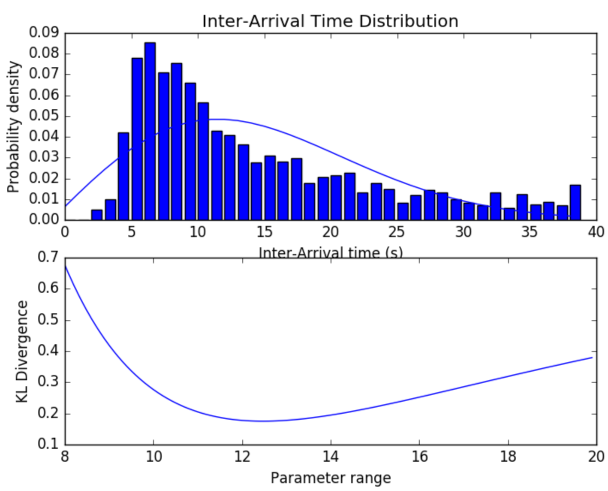

Fig. 16. Plot1) PDF for Simulated and Best-fit Distribution
Plot2) KL Divergence vs Sigma(parameter) plot which gives minima at sigma = 12.47

The Rayleigh parameter, after applying our algorithm, for the best-fit distribution (Sigma = 12.47)

B. Inter Arrival Time with Exponential

The Exponential parameter, after applying our algorithm, for the best-fit distribution (Lambda = 0.043)

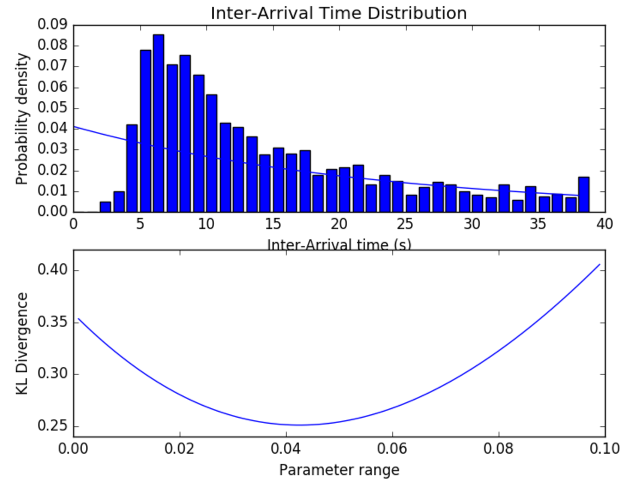

Fig. 17. Plot1) PDF for Simulated and Best-fit Distribution
Plot2) KL Divergence vs Lambda(parameter) plot which gives minima at lambda = 0.043

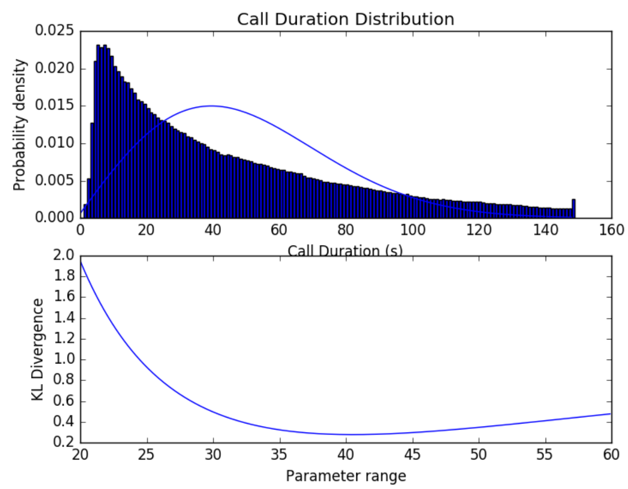

Fig. 18. Plot1) PDF for Simulated and Best-fit Distribution
Plot2) KL Divergence vs Lambda(parameter) plot which gives minima at lambda = 0.018

C. Call Duration Distribution with Rayleigh

The Rayleigh parameter, after applying our algorithm, for the best-fit distribution (Sigma = 12.47)

The Exponential parameter, after applying our algorithm, for the best-fit distribution (Lambda = 0.018)

VII. Applications

A. Cognitive Radio Sensing

We know that the spectrum we have is a limited natural resource which ought to be used carefully, efficiently and intelligently. The demand for radio spectrum like always is on an increase. The measurement results show that spectrum is not used wisely, a lot of channels are unoccupied whereas some channels are overly crowded. Cognitive radio sensing is a

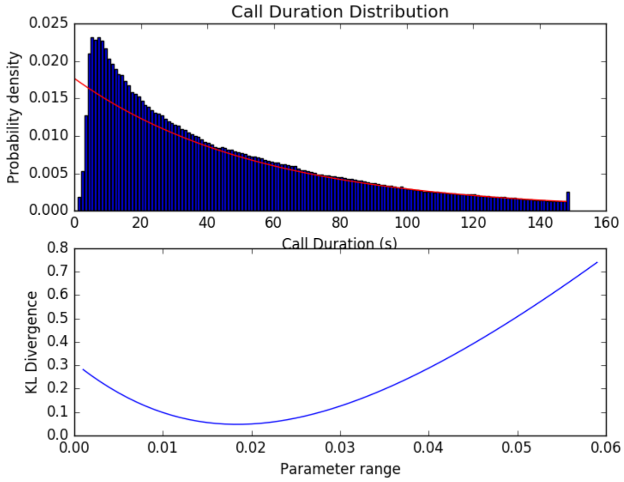

Fig. 19. Plot1) PDF for Simulated and Best-fit Distribution
Plot2) KL Divergence vs Lambda(parameter) plot which gives minima at sigma = 0.018

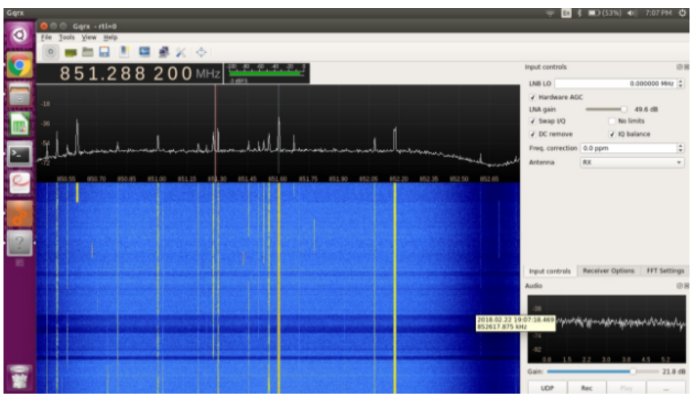

Fig. 20. Spectrogram while recording data

technique used mainly in wireless systems to occupy channels which are not in use and not the channels which are already being used in order to reduce or eradicate interference and congestion of any kind. Using our algorithm, we plan to check whether a channel is occupied or unoccupied using sweeping techniques to allot unoccupied channels to new users to avoid interference.

B. Spectrum Security and Management

By studying the spectrum usage and efficiently fitting it to a probability distribution, we can predict the usage of the spectrum at different times by different parties. While reading and analysing the spectrum usage in real-time, we can detect anomalies when the usage diverts from the probability distribution. This could be due to some parties illegally accessing the spectrum. Therefore this project is an excellent security measure for spectrum usage.

REFERENCES

[1] https://en.wikipedia.org/wiki/Poisson$_{d}istribution$
[2] https://en.wikipedia.org/wiki/Software-defined$_{r}adio$
[3] https://www.rtl-sdr.com/about-rtl-sdr/
[4] https://en.wikipedia.org/wiki/Cognitive$_{r}adio$
[5] https://en.wikipedia.org/wiki/Kullback%E2%80%93 Leibler$_{d}ivergence$
[6] https://www.countbayesie.com/blog/2017/5/9/kullback-leibler-divergence-explained
[7] https://en.wikipedia.org/wiki/Rayleigh$_{d}istribution$
[8] https://en.wikipedia.org/wiki/Normal$_{d}istribution$
[9] https://en.wikipedia.org/wiki/Exponential$_{d}istribution$
[10] http://www.statisticshowto.com/rician-distribution-rice/
[11] https://en.wikipedia.org/wiki/Rice$_{d}istribution$
[12] https://brilliant.org/wiki/poisson-distribution/