Caltech 40 metre Interferometer

Detector Characterisation

SPECTRAL PROPERTIES OF THE DATA

Department of Physics
The Australian National University (ANU)

Sub-group of ACIGA

Susan Scott
David McClelland
Bernard Whiting
Philip Charlton
Benjamin Evans
John Sandeman

Data analysis sub-group came into existence six months ago

Preliminaries

- Installed and ran GRASP on several different SGI operating systems
- Read all available 40 metre data into ANU structured Mass Data Store
- Provision of interfaces between Matlab, FRAME and GRASP codes

Problem of Out of Lock Data

Our analysis requires the use of "good data"

i.e. data taken when the instrument is in lock

Use lock channel IFO_Lock?

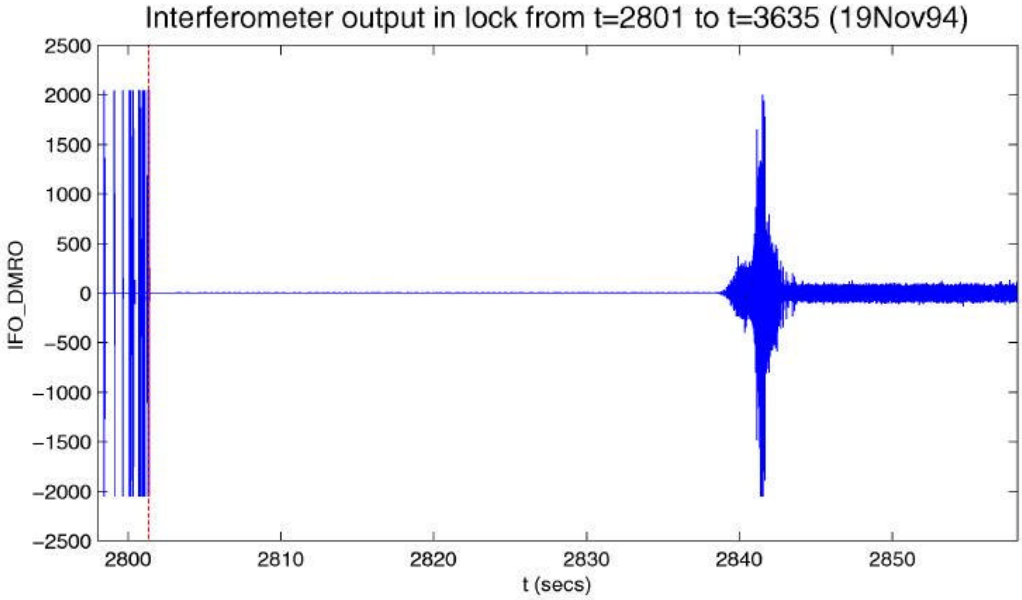
Yes, but it's too coarse for our purposes

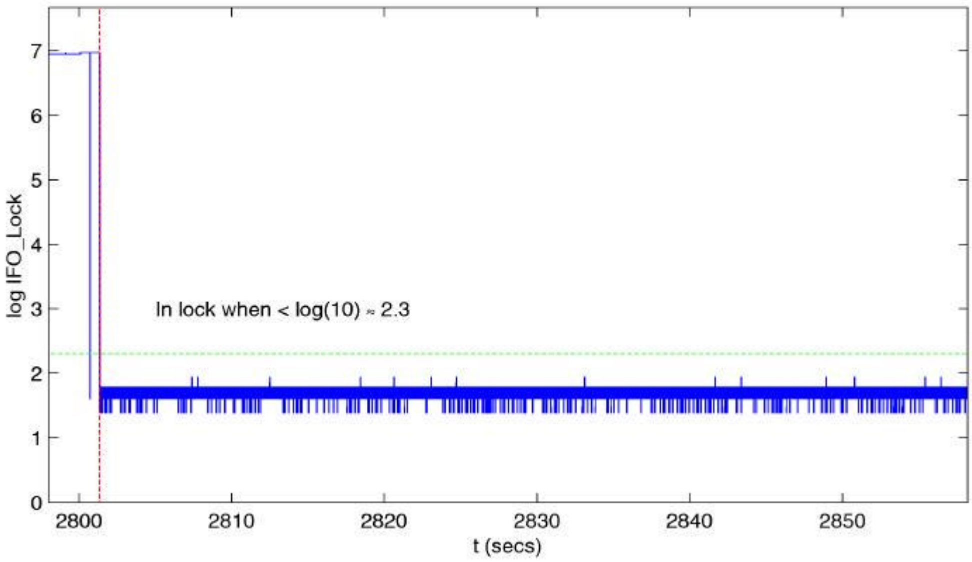
P R O B L E M : the interferometer is not actually in lock

for all sections of the lock channel showing "in lock data"

Manual fix : eliminate bad data by inspection

How could this process be carried out automatically?


Interferometer output in lock from t=2801 to t=3635 (19Nov94)
IFO_DMRO
t (secs)


log IFO_Lock
In lock when < log(10) ≈ 2.3
t (secs)

Automatic Procedure

Step 1 : Only consider data which the lock channel indicates is "in lock"

Step 2 : Segment this "in lock" data into blocks of equal length e.g. 1,000 data points

Step 3 : Compute the mean signal $\mathbf{m_i}$ for each block and the associated standard deviation $\mathbf{s_i}$ for each block

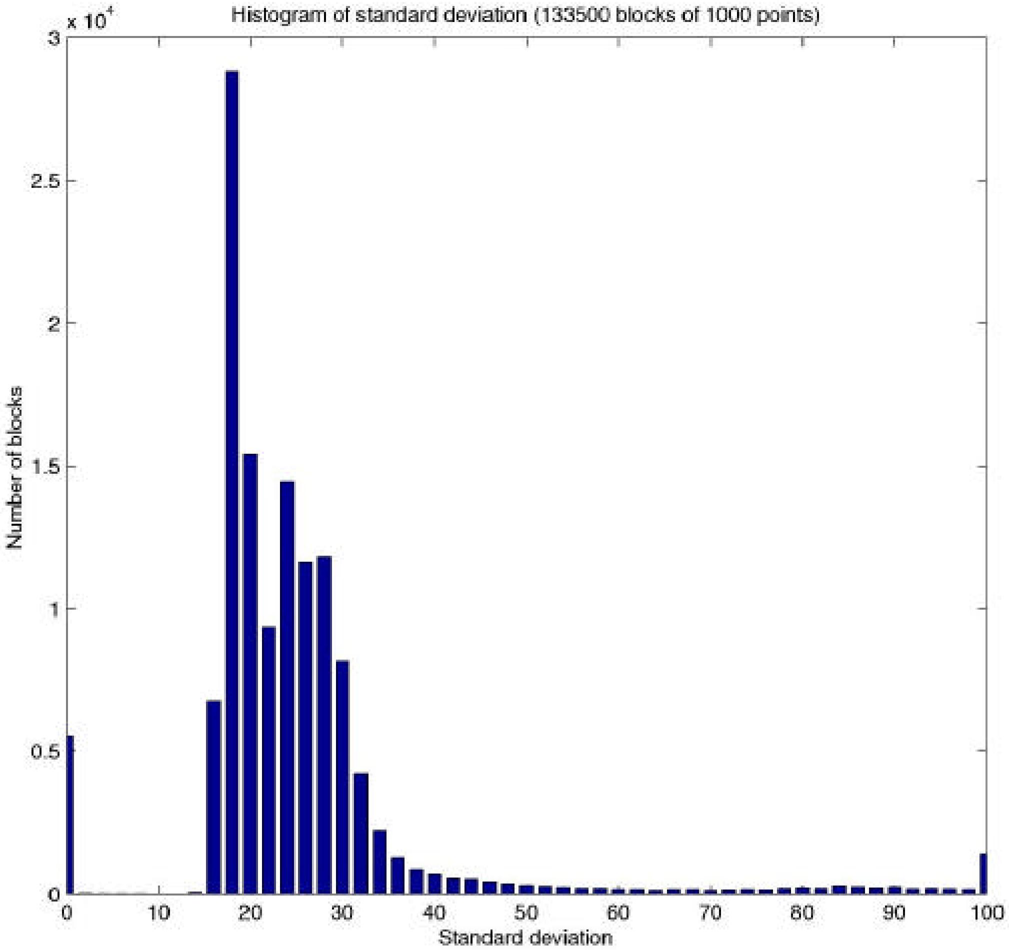
Step 4 : Histogram the standard deviations $\mathbf{s_i}$ for all the blocks

Step 5 : Using the histogram, select suitable (non-zero) minimum and maximum standard deviations

Step 6 : Blocks with standard deviation lying between the minimum and maximum are marked as **good**

Step 7 : The remaining blocks are marked as **bad** and discarded

Objective: to deliver code for this procedure to LIGO by June 2000



Histogram of standard deviation (133500 blocks of 1000 points)
x 10^4
3
2.5
2
1.5
1
0.5
0
Number of blocks
0
10
20
30
40
50
60
70
80
90
100
Standard deviation

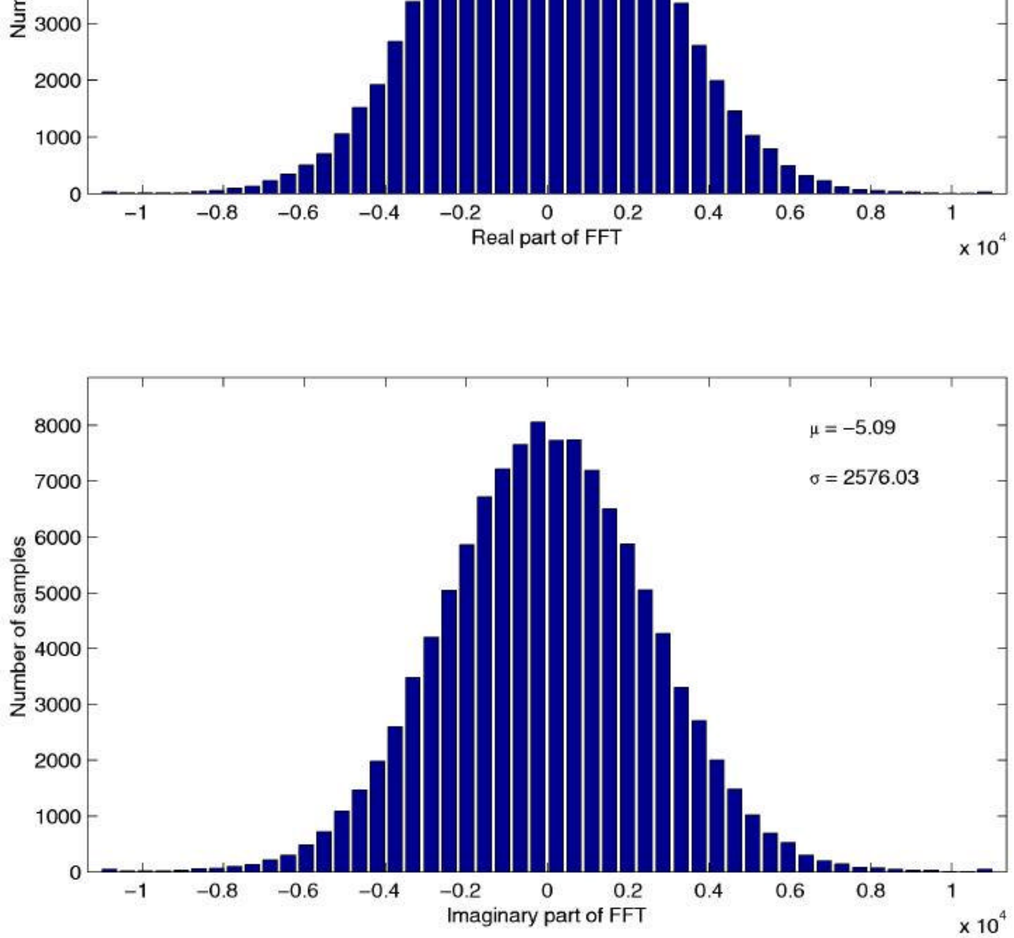
F requency Histograms

- Use **good** blocks of 1,000 data points
- Pad the blocks out to 8,192 = 2^{13} points by adding zeros
- Consider the first 550 blocks of data
- Perform a Fast Fourier Transform (FFT) on each block
- Compile the FFT data into a 550×4097 array
- For each frequency:
  (a) **histogram** the real part of the FFT (50 bins)
  calculate the mean and standard deviation of the sample
  (b) **histogram** the imaginary part of the FFT (50 bins)
  calculate the mean and standard deviation of the sample

- Execute the same procedure for the next 550 blocks of data
- Combine with the first 550 block to compile histograms:

$$m \rightarrow \frac{N}{N+M} m + \frac{M}{N+M} m_{\text{new}}$$

$$s^2 \rightarrow \frac{N}{N+M} s^2 + \frac{M}{N+M} s^2_{\text{new}} + \frac{NM}{(N+M)^2} (m - m_{\text{new}})^2$$


Histograms at 599.91 Hz (f_n=499) from 114617000 points
μ = −5.75
σ = 2574.59
Number of samples
Real part of FFT
x 10^4
μ = −5.09
σ = 2576.03
Number of samples
Imaginary part of FFT
x 10^4

Likelihood Ratio as a Measure of Gaussianity

- Our samples of the real and imaginary components of the FFT for a given frequency are binned
- We can therefore assume that they are distributed **multinomially**, with probability $\boldsymbol{p_i}$ of a point falling into bin $\boldsymbol{i}$

The ***likelihood function L*** for a multinomial distribution is given by

$$L = \frac{N!}{n_1!n_2!\ldots n_k!}\, p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k}$$

$\boldsymbol{k}$ is the number of bins

$\boldsymbol{n_i}$ is the number of points in bin $\boldsymbol{i}$

$N = \sum_{i=1}^{k} n_i$ is the total number of points

A ***likelihood ratio l*** is obtained by taking the ratio of ***L*** to the **maximum value** attainable by ***L*** as the $\boldsymbol{n_i}$'s vary

N.B. subject to the constraint $N = \sum_{i=1}^{k} n_i$

This gives a value $0 \le l \le 1$

Example

For ***2M*** tosses of an unbiased coin, the likelihood of ***n*** heads is

$$\frac{(2M)!}{n!(2M-n)!}\left(\frac{1}{2}\right)^{n}\left(\frac{1}{2}\right)^{2M-n}$$

The **maximum likelihood** is attained when ***n = M***. Hence

$$l = \frac{M!M!}{n!(2M-n)!}\left(\frac{1}{2}\right)^{n-M}\left(\frac{1}{2}\right)^{(2M-n)-M}$$

$$= \frac{(M!)^2}{n!(2M-n)!}$$

The likelihood ratio is used to measure the Gaussianity of binned data

- Take the probabilities p_i from a **normal distribution** with mean **m** and standard deviation **s** :

$$p_i = \int_{x_i-\Delta x/2}^{x_i+\Delta x/2} \frac{1}{s\sqrt{2p}} \exp\left(-\frac{1}{2}\left(\frac{x-m}{s}\right)^2\right)dx$$

x_i is the centre of the ***i*** th bin

D*x* is the bin width

m and **s** are obtained from the frequency data

L attains its **maximum** at $\hat{n}_i = Np_i$

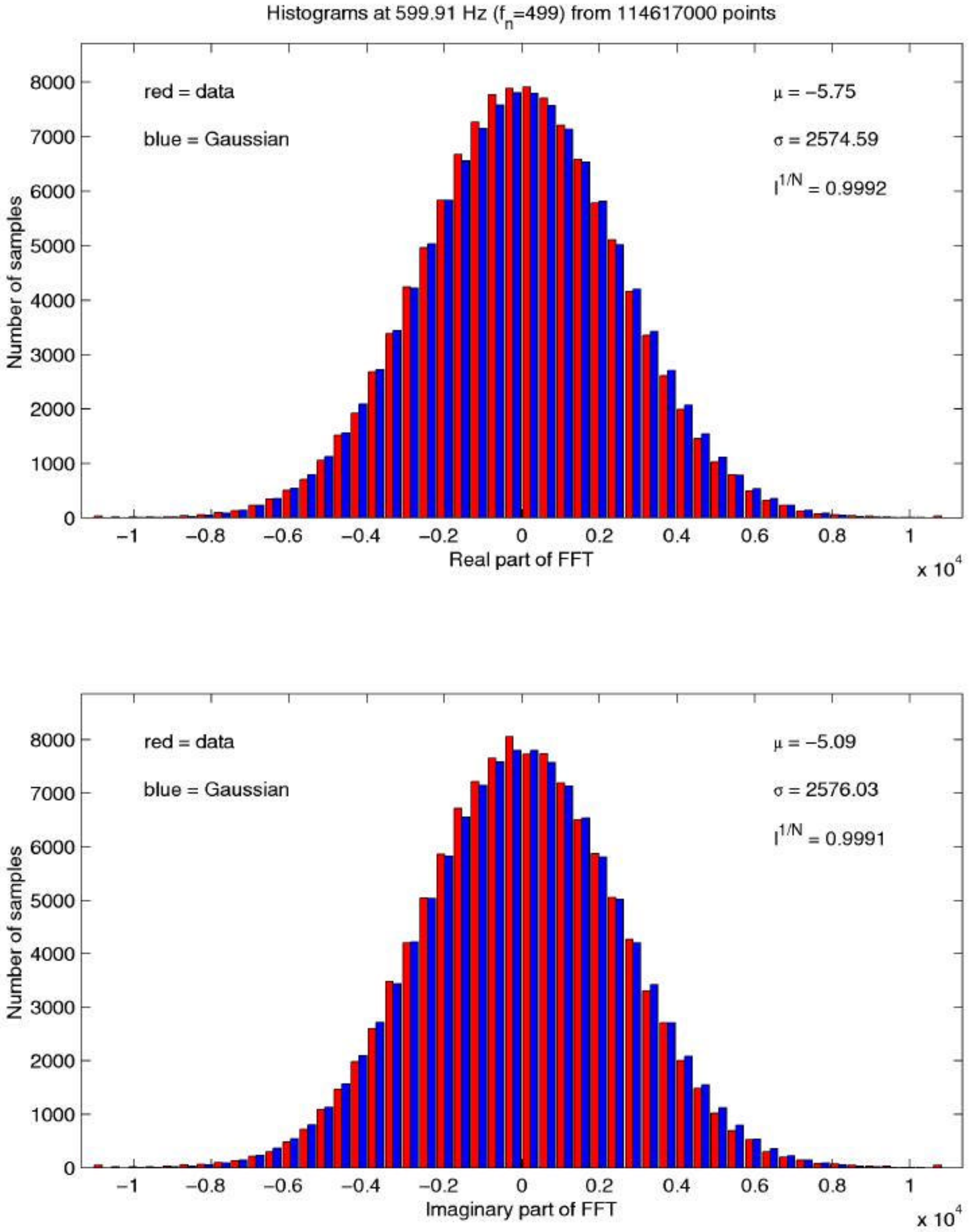
Thus ***l*** is given by

$$l = \frac{\hat{n}_1!\hat{n}_2!\ldots\hat{n}_k!}{n_1!n_2!\ldots n_k!} p_1^{n_1-\hat{n}_1} p_2^{n_2-\hat{n}_2} \ldots p_k^{n_k-\hat{n}_k}$$

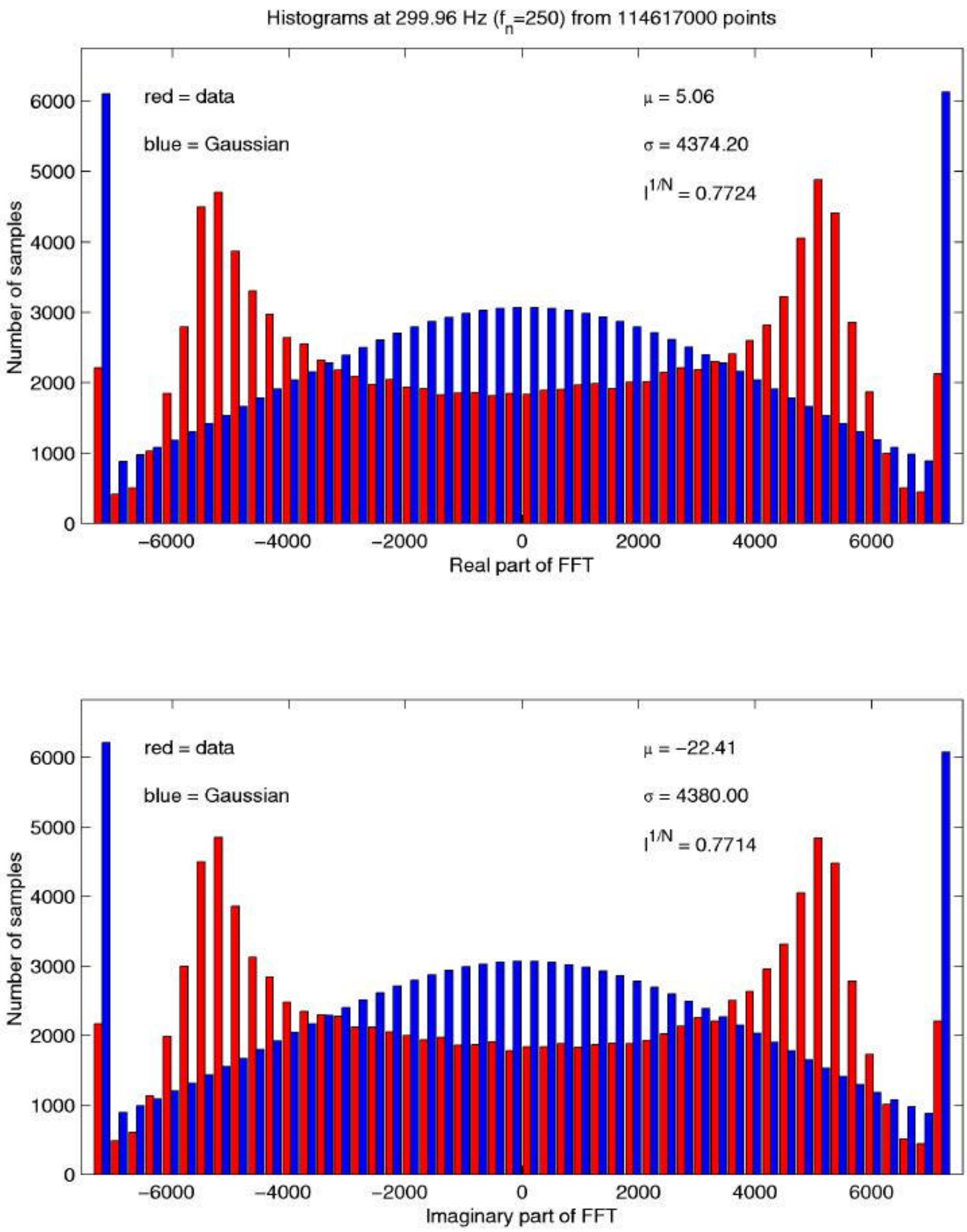
In practice it is simpler to calculate log $\boldsymbol{l}$

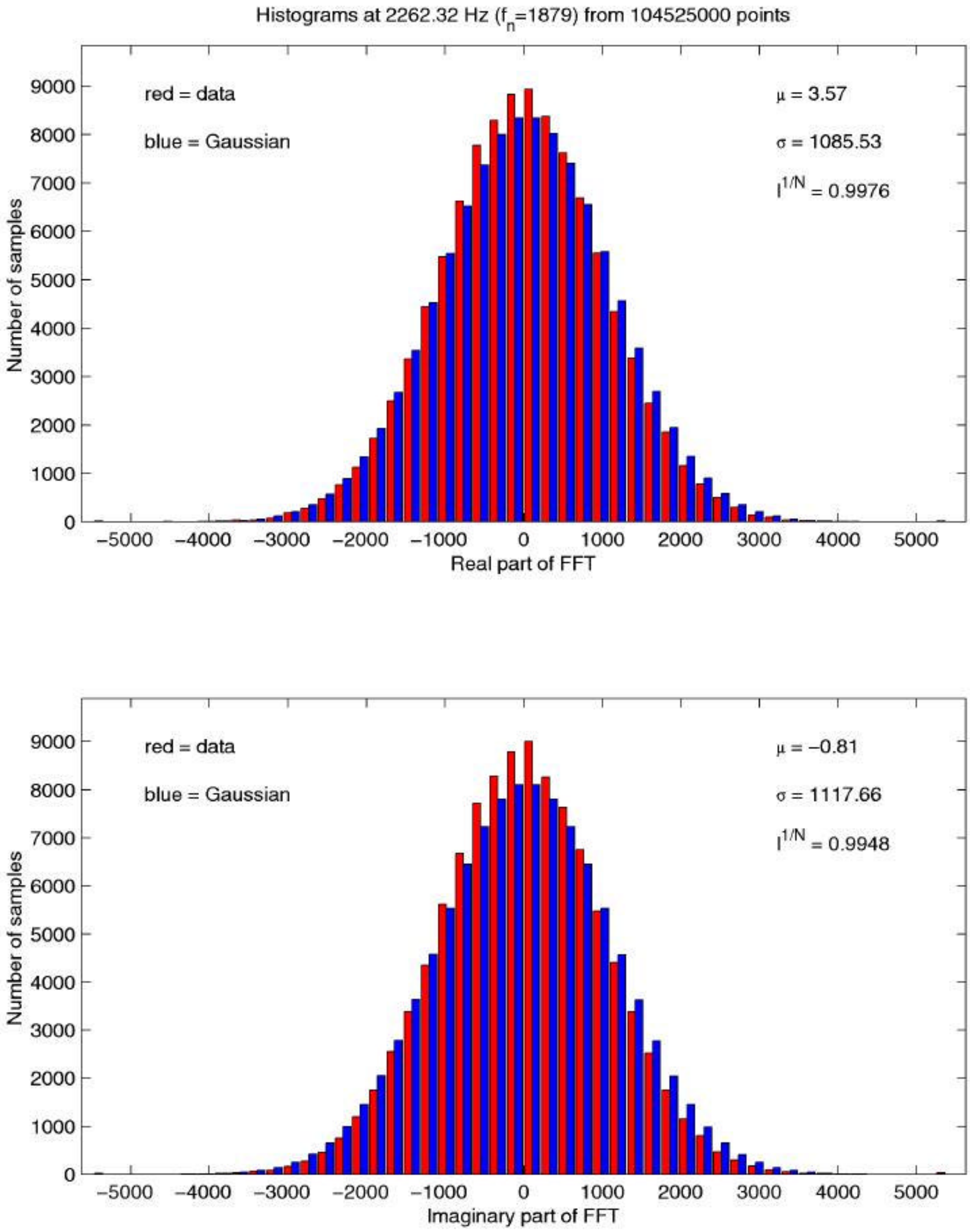
- Values of $\boldsymbol{l}$ close to 1 indicate a good fit to Gaussian
- For large $\boldsymbol{N}$, $\boldsymbol{l}$ is extremely sharply peaked
- A more useful measure of **Gaussianity** is given by $\boldsymbol{l^{1/N}}$

  the $\boldsymbol{N}$ th root of $\boldsymbol{l}$

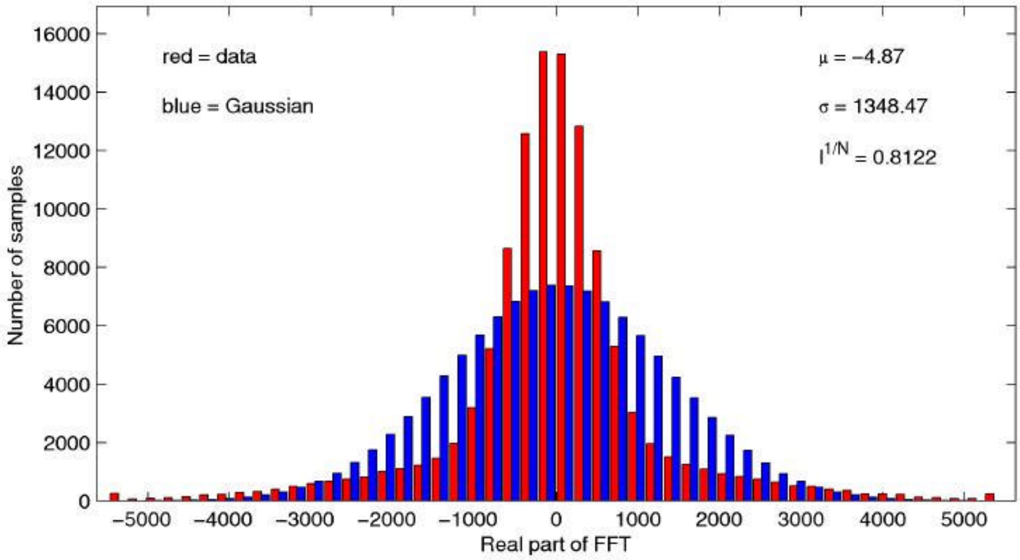


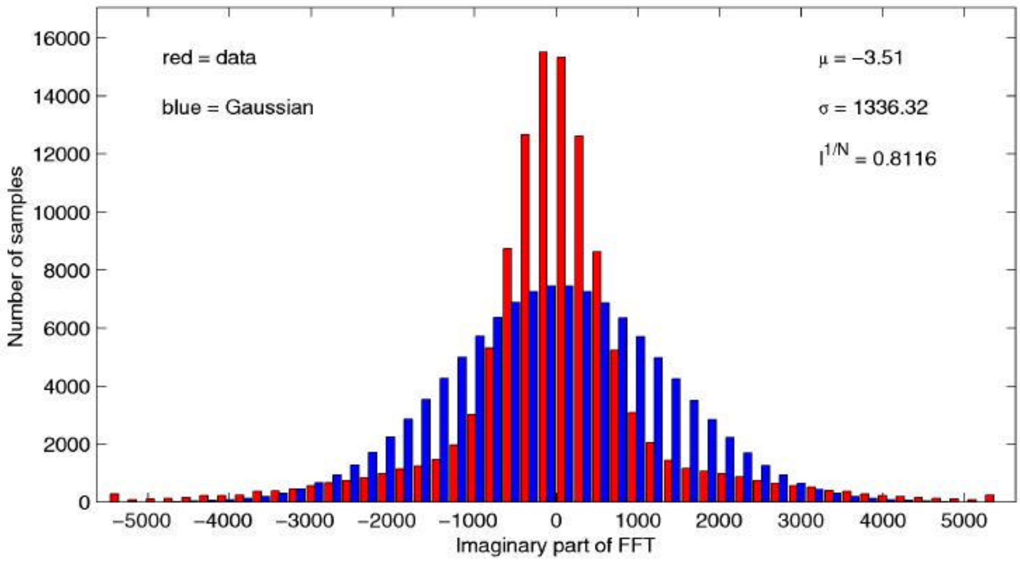
Histograms at 599.91 Hz (f_n=499) from 114617000 points
red = data
blue = Gaussian
μ = −5.75
σ = 2574.59
l^1/N = 0.9992
Number of samples
Real part of FFT
x 10^4
red = data
blue = Gaussian
μ = −5.09
σ = 2576.03
l^1/N = 0.9991
Number of samples
Imaginary part of FFT
x 10^4


Histograms at 299.96 Hz (f_n=250) from 114617000 points
red = data
blue = Gaussian
μ = 5.06
σ = 4374.20
$I^{1/N}$ = 0.7724
Number of samples
Real part of FFT
red = data
blue = Gaussian
μ = −22.41
σ = 4380.00
$I^{1/N}$ = 0.7714
Number of samples
Imaginary part of FFT


Histograms at 2262.32 Hz (f_n=1879) from 104525000 points
red = data
blue = Gaussian
μ = 3.57
σ = 1085.53
I^{1/N} = 0.9976
Number of samples
Real part of FFT
red = data
blue = Gaussian
μ = −0.81
σ = 1117.66
I^{1/N} = 0.9948
Number of samples
Imaginary part of FFT

Histograms at 2262.32 Hz (f_n=1879) from 114617000 points





Work in Progress

- Speed tests on the SGI PowerChallenge

- Further measures of normality

  e.g. skew, kurtosis, χ^2-statistic

- Apply line removal techniques to 40 metre data

- Use the statistical analysis techniques outlined to measure the effects of different line removal techniques

  (40 metre, Glasgow)