Investigating Higher Order Statistics and Gaussianity

Steve Penn, Syracuse University

LIGO-G010132-00-Z

Synopsis

- Introduction to Higher Order Statistics
  - 1D: Correlation, Coherence, Power Spectra
  - 2D: Bicorrelation, Bicoherence, Bispectrum
  - 3D…
- Bispectrum diagnostic
- Gaussianity Test
- Linearity Test

What are Higher Order Statistics?

- 1D Statistics:
  - » Correlation: $C_{xy}(t) = \int_{-\infty}^{\infty} x(\tau) y(t+\tau)\, d\tau \quad \Leftrightarrow \quad X(f)\, Y^{*}(f) = S_{xy}(f)$
  - » Power Spectral Density: $C_{2x}(t) \quad \Leftrightarrow \quad X(f)\, X^{*}(f) = S_{2x}(f)$
  - » Coherence: $C_{xy}(f) = \dfrac{S_{xy}(f)}{\sqrt{S_{2x}(f)\, S_{2y}(f)}}$
    - – Tells us power and phase coherence at a given frequency

Second Order Statistics

- 2D Statistics:
  - Cumulant:

$$C_{xyz}(t,t') = \int_{-\infty}^{\infty} x(\tau)\, y(t+\tau)\, z(t'+\tau)\, d\tau \quad \Leftrightarrow \quad X(f_1)\, Y(f_2)\, Z^*(f_1+f_2) = S_{xyz}(f_1, f_2)$$

  - Bispectral Density:

$$C_{3x}(t) \quad \Leftrightarrow \quad X(f_1)\, X(f_2)\, X^*(f_1+f_2) = S_{3x}(f_1, f_2)$$

  - Bicoherence:

$$\mathrm{C}_{xyz}(f) = \frac{S_{xyz}(f_1, f_2)}{\sqrt{S_{2x}(f_1)\, S_{2y}(f_2)\, S_{2z}(f_1, f_2)}}$$

    - Tells us power and phase coherence at a coupled frequency

Zero-lag Cumulants

Mean	Variance	Skewness	Kurtosis
$C_x(0)$	$C_{2x}(0)$	$C_{3x}(0)$	$C_{4x}(0)$
		0 if Symmetric	0 if Gaussian

Useful statictical values, but…

Skewness = 0 does not prove symmetry

Kurtosis = 0 does not prove Gaussianity

Variations in skew and kurtosis not well quantified.

Why Higher Order Statistics?

- For a Gaussian process: $C_{nx}(t) = 0, \text{ for } n > 2$

- For independent processes:

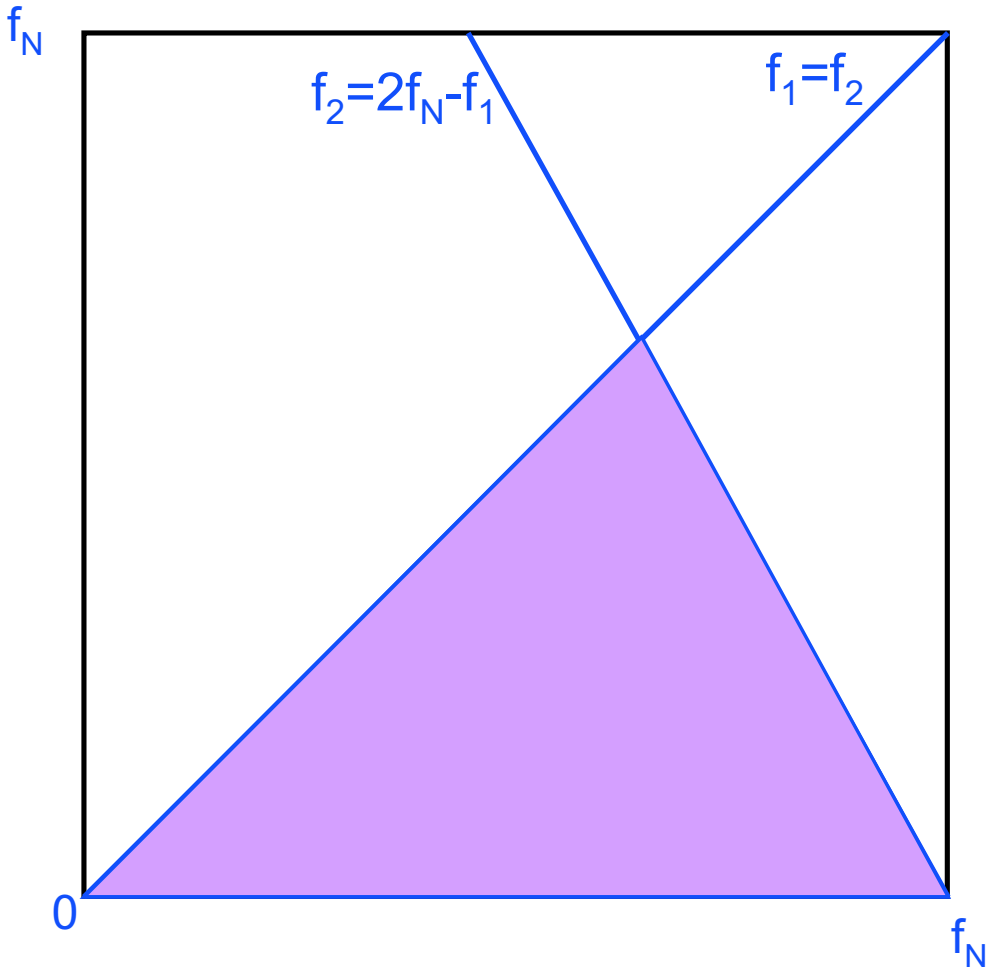
$$z(t) = x(t) + y(t), \quad C_{nz}(t) = C_{nx}(t) + C_{ny}(t) \xrightarrow{n>2} C_{ny}(t)$$

- Allows for separation of Gaussian process for n>2
  - » Visual check of frequency coupling and phase noise
  - » Statistical test for the probability of gaussianity and linearity
  - » Iterative process to reconstruct nongaussian signal from the higher order cumulants

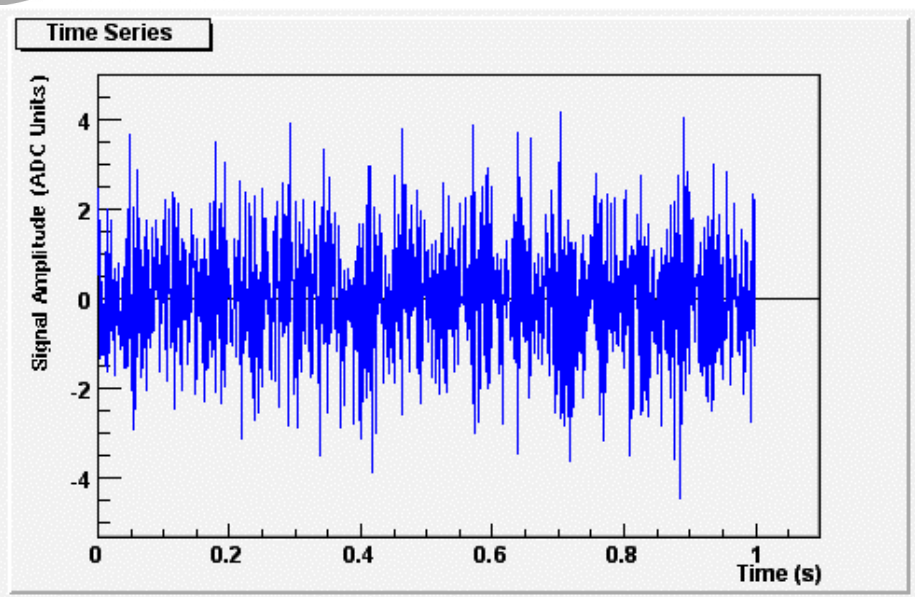
Gaussianity Monitor

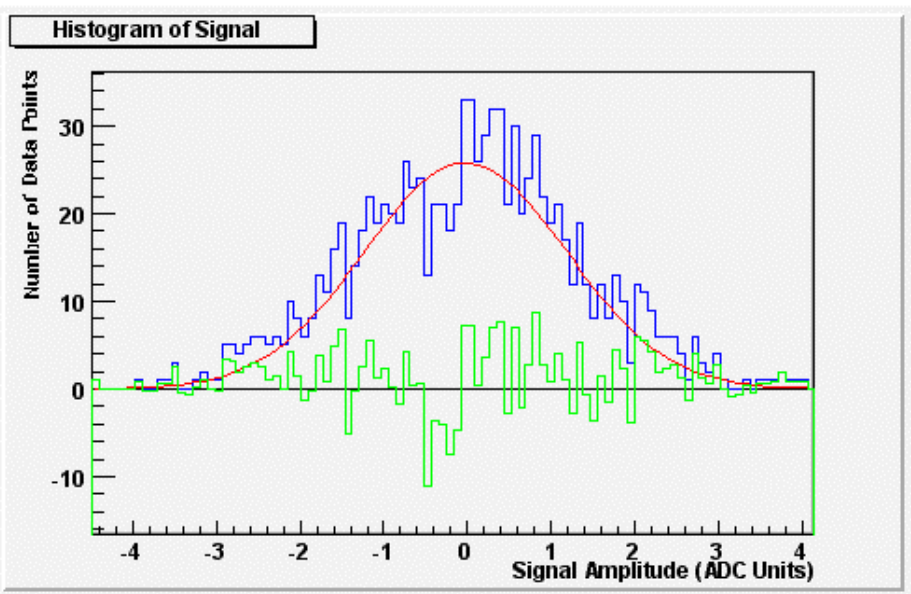
- Diagnostic Plot:
  - » Time series, Power spectrum, Two perspectives of the histogram of time series with gaussian fit
- Bispectrum and Bicoherence Plot
- Gaussianity test: $\int C_{3x}(f_1, f_2)\, d\Omega_U = \chi^2_{2D}$
- Summary file:
  - » Frame, channel, mean, variance, skew, kurtosis, gaussianity probability

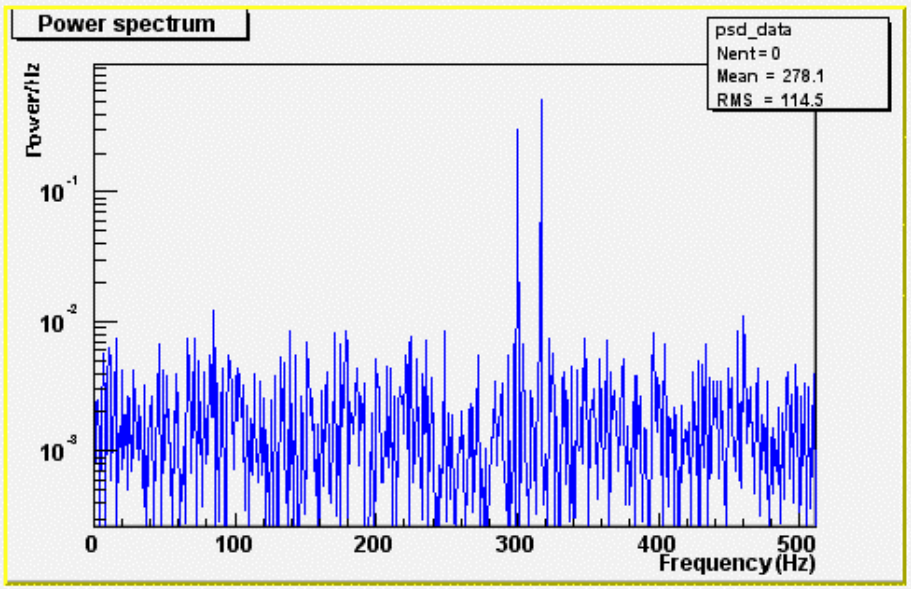
Bispectrum Unique Area

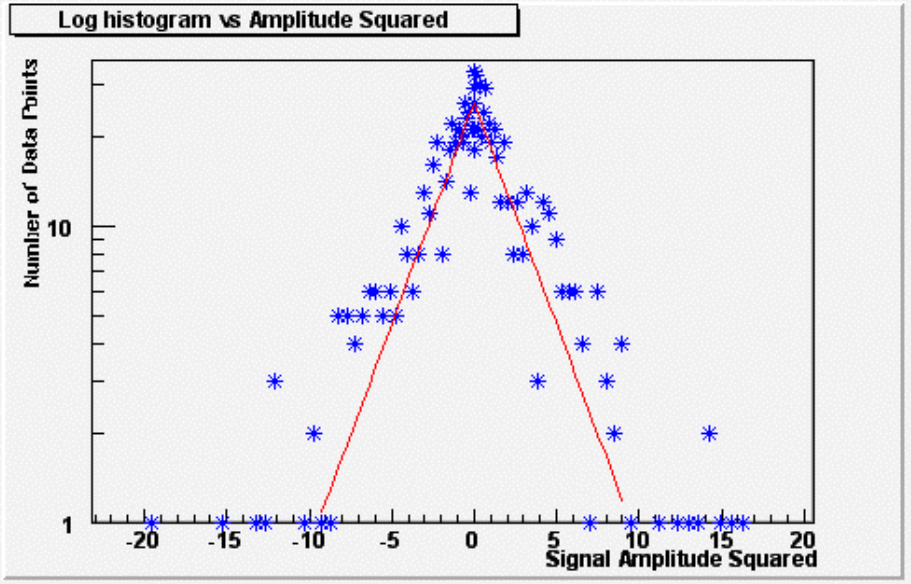


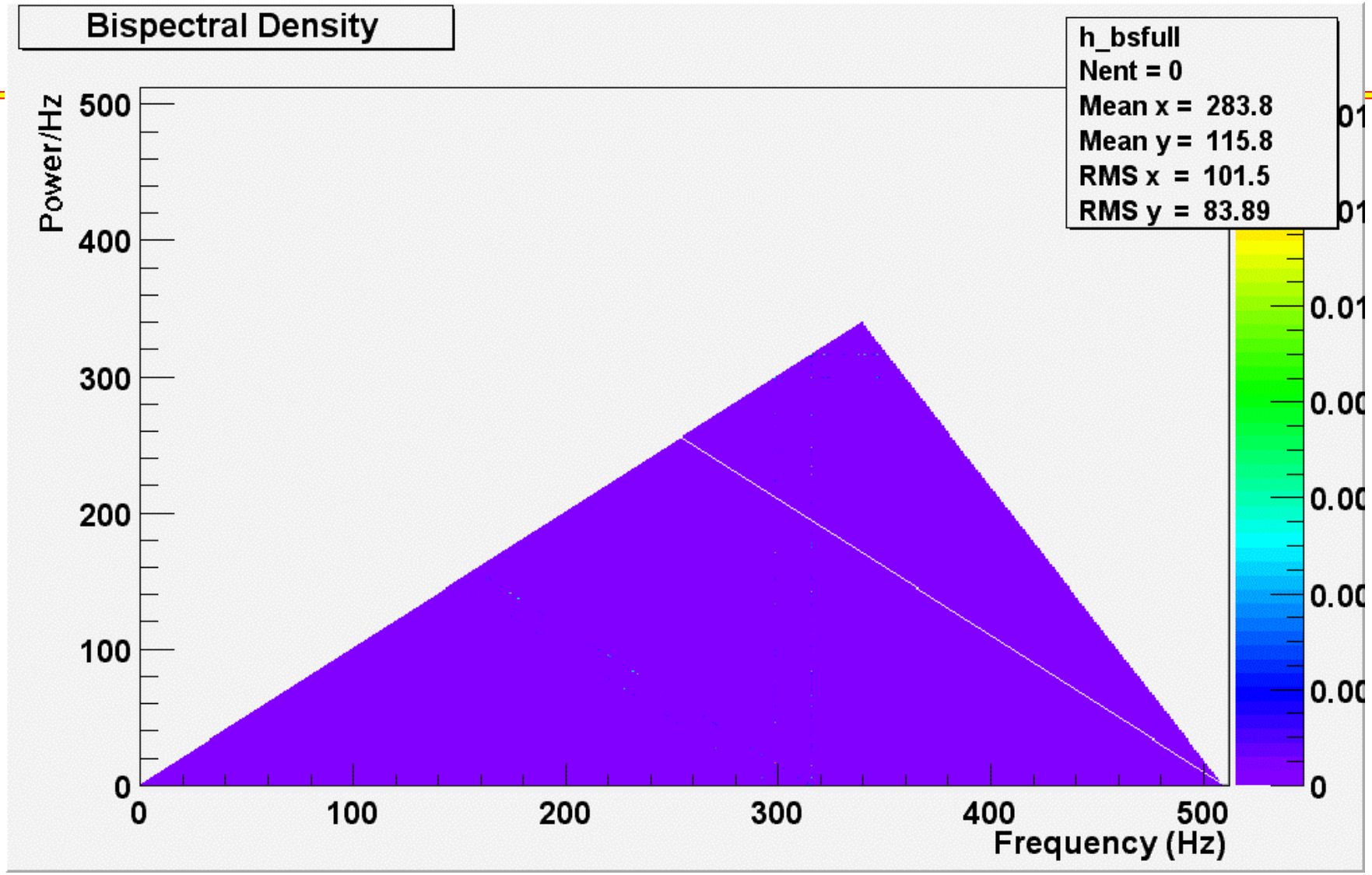
LIGO

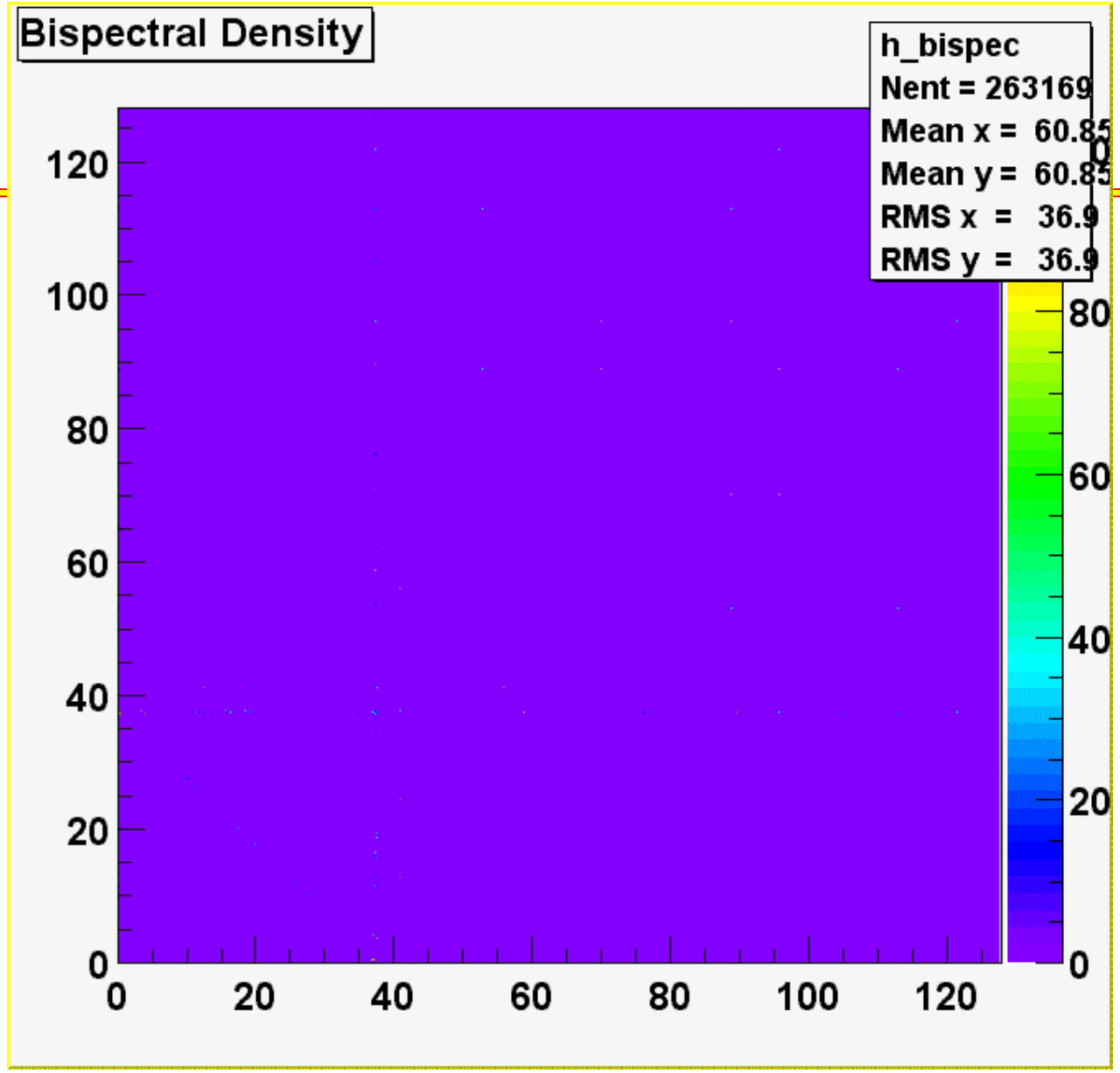

Time Series
Signal Amplitude (ADC Units)
Time (s)


Histogram of Signal
Number of Data Points
Signal Amplitude (ADC Units)


Power spectrum
psd_data
Nent = 0
Mean = 278.1
RMS = 114.5
Power/Hz
Frequency (Hz)

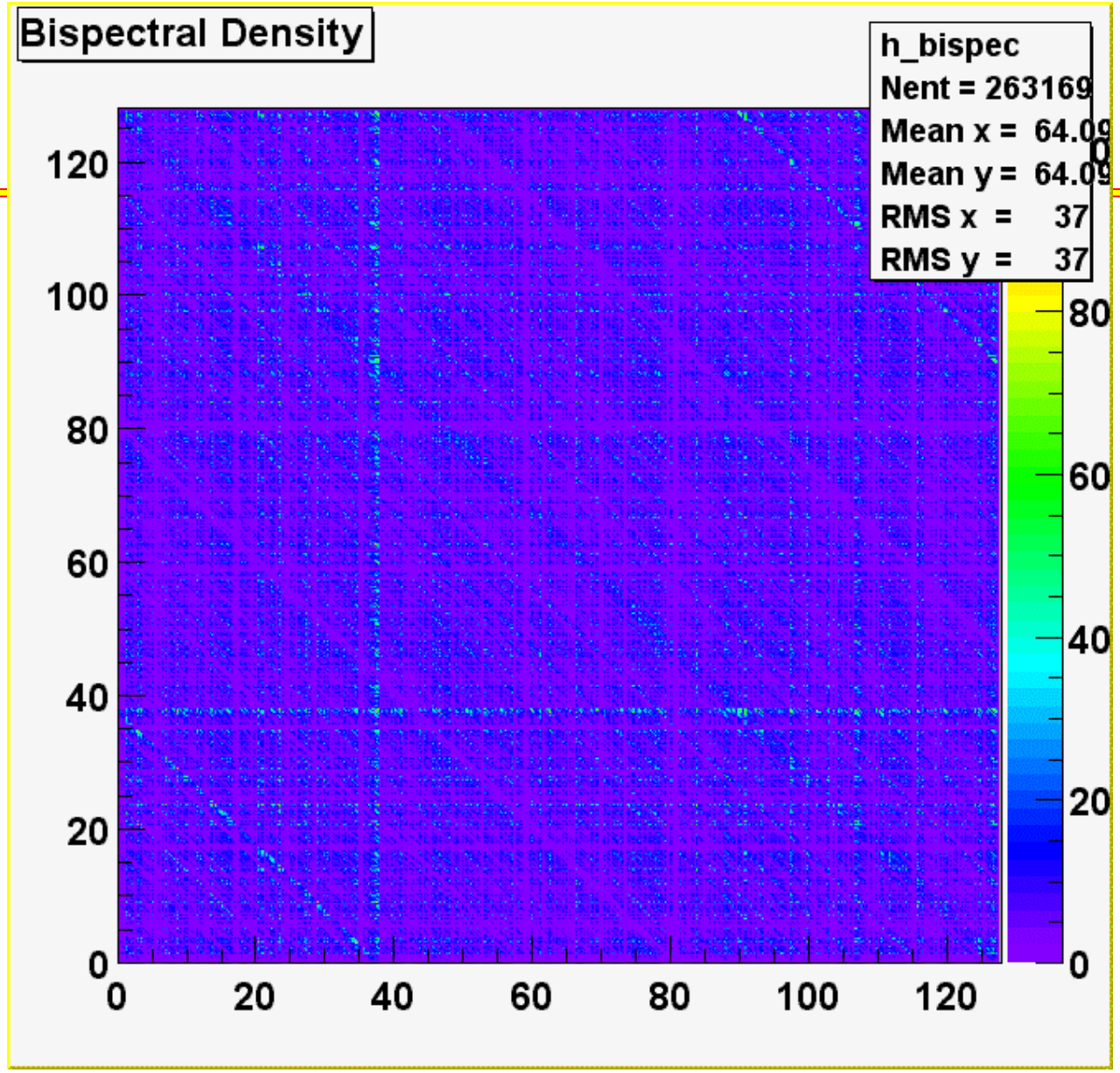

Log histogram vs Amplitude Squared
Number of Data Points
Signal Amplitude Squared


Bispectral Density
h_bsfull
Nent = 0
Mean x = 283.8
Mean y = 115.8
RMS x = 101.5
RMS y = 83.89
Power/Hz
0
100
200
300
400
500
Frequency (Hz)


Bispectral Density
h_bispec
Nent = 263169
Mean x = 60.85
Mean y = 60.85
RMS x = 36.9
RMS y = 36.9
0
20
40
60
80
100
120

Status and Conclusions

- Using the HOSA package in Matlab was not practical. We needed a monitor running in DMT.
- Old Bispectral algorithm *extremely* slow (100*realtime)
- New algorithm using FFTW appears to be near realtime
- Now we can analyze some data!!
- Upcoming additions:
  - » Better windowing
  - » User selectable time interval
  - » Record # events above user selected σ limit
  - » Plot/Set reference distribution for each channel.


Bispectral Density
h_bispec
Nent = 263169
Mean x = 64.09
Mean y = 64.09
RMS x = 37
RMS y = 37
0
20
40
60
80
100
120