Event Weighted Tests for Detecting Periodicity in Photon Arrival Times

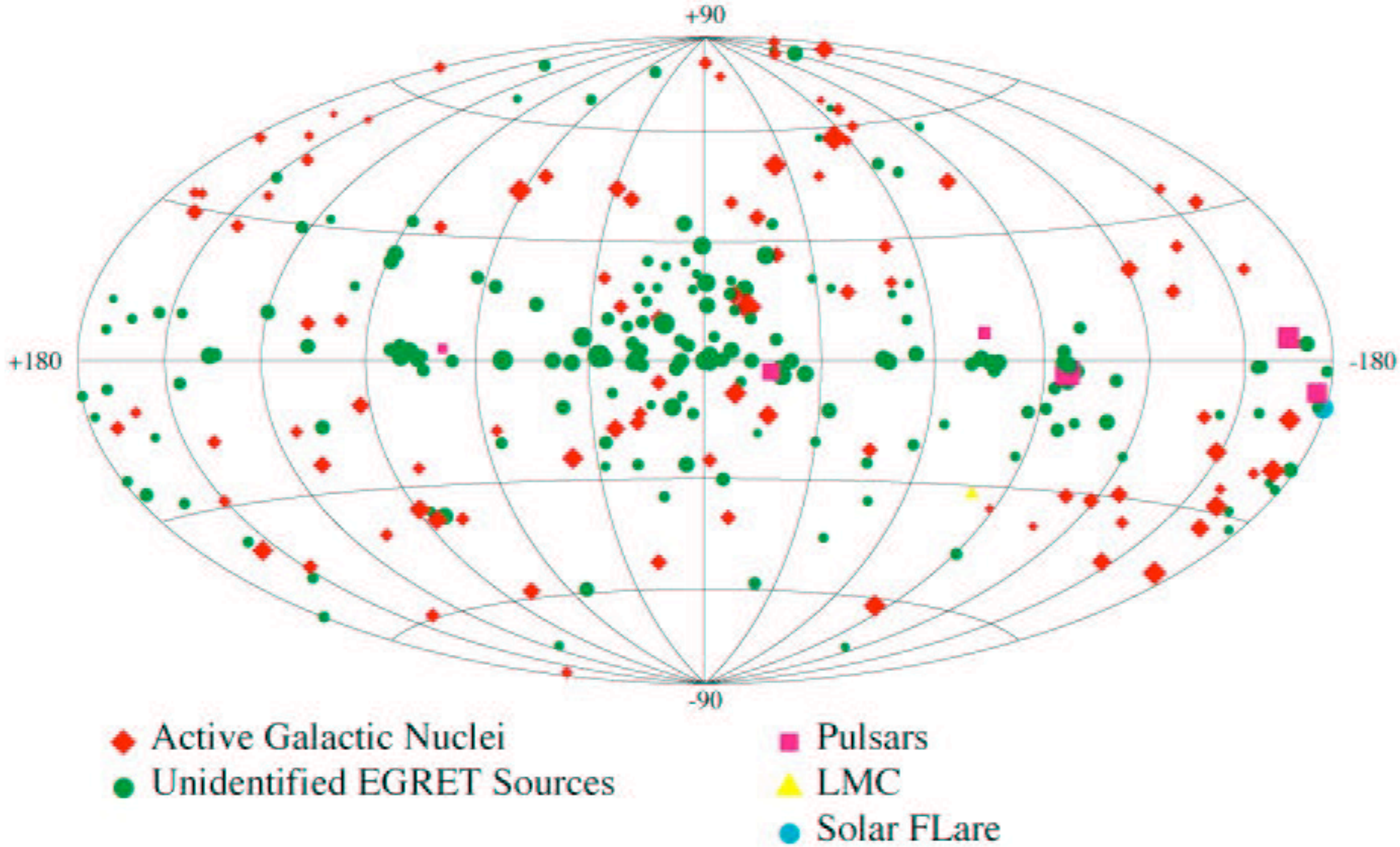
John Rice
Statistics Department
University of California, Berkeley

Joint work with Peter Bickel and Bas Kleijn. Thanks to Seth Digel, Patrick Nolan
Tom Loredo, Charlotte Wickham, Jeremy Shen

Outline

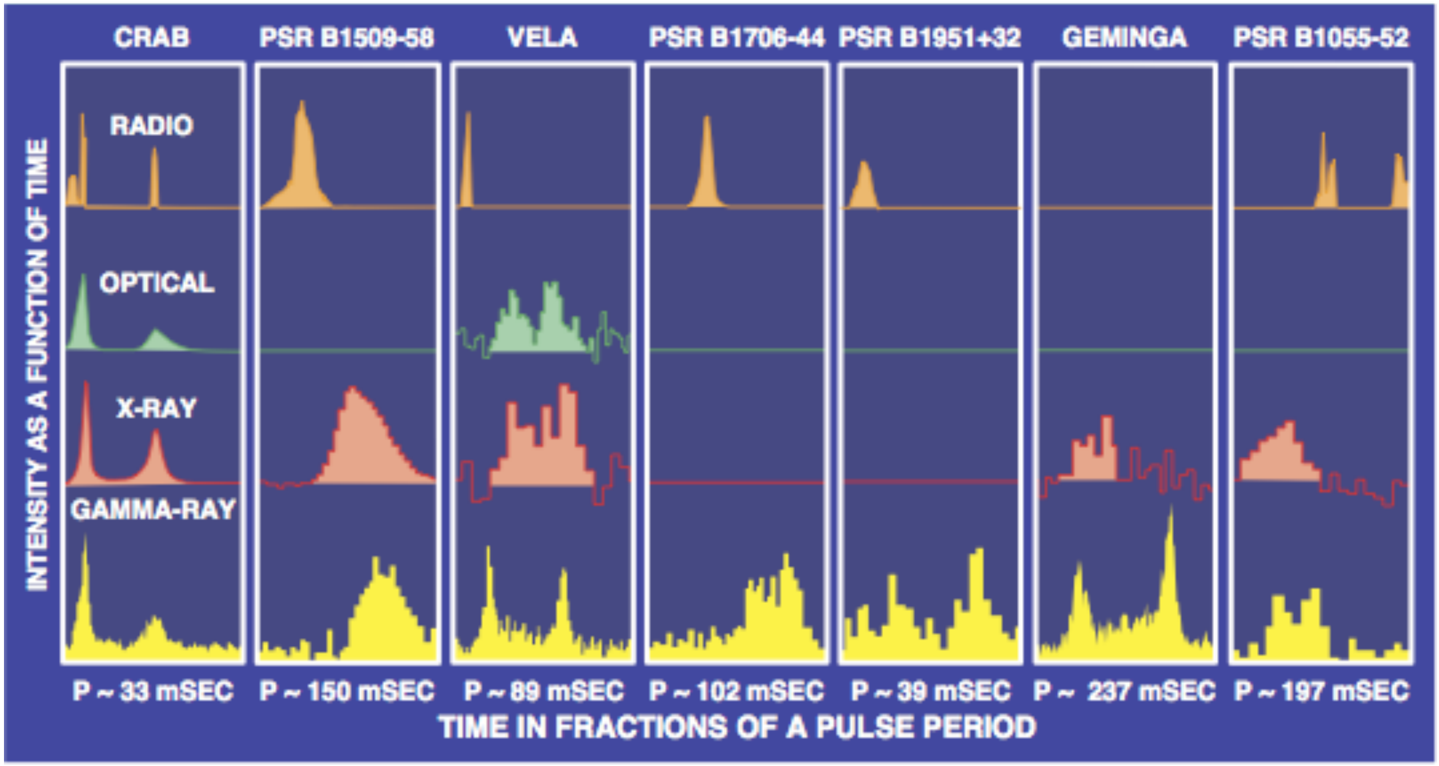
- Motivation: EGRET sources and gamma-ray pulsars
- Detection as an hypothesis testing problem
- A score test and discussion of its properties
- Difficulties of a blind search
- Integration over frequency bands as an alternative to discretization
- Use of simulation and extreme value theory to assess significance

EGRET Sources



Gamma-Ray Pulsars

radio quiet



Many gamma-ray sources are unidentified and may be pulsars, but establishing that these sources are periodic is difficult. Might only collect ~1500 photons during a 10 day period.

Detection problem

Barycentric corrected arrival times $0 < t_1 < \ldots < t_N < T$. For EGRET, N ~ 1000, T ~ 10 days

Energy and incidence angle of each photon. *How should this information be used?*

Some photons come from source and some from background.

Question: Is the source periodic, perhaps with decaying frequency? Pulse profile unknown.

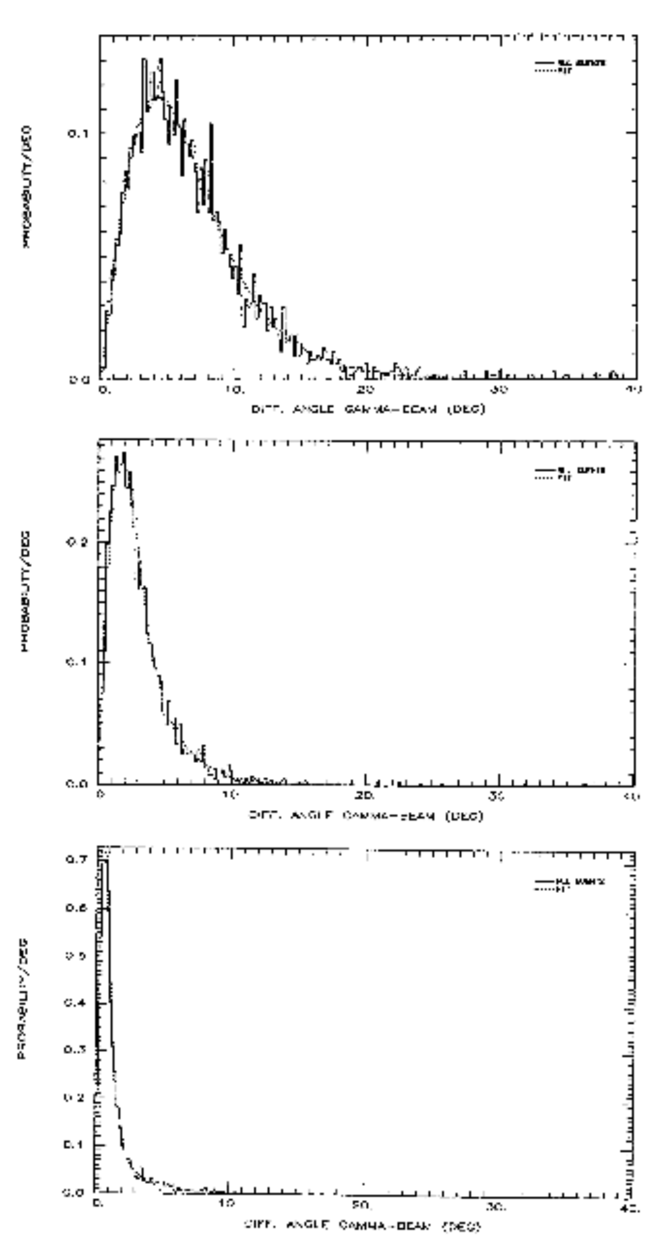


FIG. 17. EGRET angular dispersion, shown in the form probability/DEG: (a) 50 MeV; (b) 200 MeV; (c) 1000 MeV

Detection problem as hypothesis test

Unpleasant fact: There is no optimal test. *Even if the frequency were known, a detection algorithm optimal for one pulse profile will not be optimal for another one. No matter how clever you are, no matter how rich the dictionary from which you adaptively compose a detection statistic, no matter how multilayered your hierarchical prior, your procedure will not be globally optimal.*

The pulse profile $\nu(t)$ is an infinite dimensional object. *Any* test can achieve high power against local alternatives for at most a *finite* number of directions. In other words, associated with any particular test is a finite dimensional collection of targets and it is only for such targets that it is highly sensitive.

Consequence: You have to be a [closet] Bayesian and choose directions *a priori*.

Lehman & Romano. *Testing Statistical Hypotheses*. Chapt 14

Specifying a target

Consider testing against a template for the pulse profile, a probability density:

$$\nu_0(t) = 1 + \eta \sum_{n \neq 0} \alpha_n e^{2\pi i n t}$$

$$\nu_\tau(t) = 1 + \eta \sum_{n \neq 0} \alpha_n e^{2\pi i n (t+\tau)}$$

Model the arrival times as a Poisson process with rate function:

$$\lambda(t|\theta, \tau, \mu, f) = \mu c(t)[(1-\theta) + \theta \nu_\tau(\phi(t))], \quad 0 \leq \theta \leq 1$$

rate   instrument sensitivity   proportion from source

$$\phi(t) = \phi_0 + ft \quad \text{or} \quad \phi(t) = \phi_0 + ft + \frac{1}{2}\dot{f}t^2$$

$H_0 : \eta = 0$

Likelihood function and score test

Associated with each event is auxiliary information, the incidence angle and the measured energy; denote these variables by z. Let $f_B(z)$ denote the probability density function of z for a background event and $f_S(z)$ the density function for a source event. The likeihood of the marked point process is

$$
\begin{aligned}
L &= \mu^N \prod_{j=1}^{N} c(t_j)[(1-\theta)f_B(z_j) + \theta f_S(z_j)\nu_\tau(\phi(t_j))] \times \\
&\quad \exp(-\mu \int_0^T c(t)[(1-\theta) + \theta\nu_\tau(\phi(t))]dt)
\end{aligned}
$$

A score test is found by evaluating the derivative of the log likelihood with respect to η at 0

Score test

Posterior probability that the photon was from the source

$$S(\tau) \;=\; \sum_{i=1}^{N}\left(\overbrace{\frac{\theta f_S(z_j)}{(1-\theta)f_B(z_j)+\theta f_S(z_j)}}^{w_j}(\nu_\tau(\phi(t)-1)\right) -$$

$$\mu\theta\int_0^T c(t)[\nu_\tau(\phi(t))-1]dt$$

Neglible for large T

To eliminate the dependence on τ, square and integrate, resulting in

$$Q_T = \frac{1}{T}\sum_{n\neq 0}|\alpha_n|^2|A_n|^2$$

where

$$A_n = \sum_j w_j \exp(2\pi i n\phi(t_j))$$

Relationship to classical tests in the un-weighted case

$$Q_T = \frac{1}{T} \sum_{n \neq 0} |\alpha_n|^2 |A_n|^2 \qquad A_n = \sum_j w_j \exp(2\pi i n \phi(t_j))$$

Rayleigh's test (1919): $\alpha_n = 0, n \geq 2$

Bucherri et al. (1983): Z_m^2: $\alpha_n = 1,\ n \leq m$

De Jager et al. (1989): choose m adaptively

Beran (1969) showed test to be locally most powerful among invariant tests for uniformity on the circle

Weight function

$$w_j = \frac{\theta f_S(z_j)}{(1-\theta) f_B(z_j) + \theta f_S(z_j)}$$

$$\begin{aligned} z &= (E, \varphi) \\ f_B(z) &= f_B(E) f_B(\varphi|E) \\ f_S(z) &= f_S(E) f_S(\varphi|E) \end{aligned}$$

$$w(E,\varphi) = \frac{\theta f_S(E) f_S(\varphi|E)}{\theta f_S(E) f_S(\varphi|E) + (1-\theta) f_B(E) f_B(\varphi|E)}$$

Depends on spectra of source and background through their ratio

Weight function

The optimal weight function depends on the ratio of background and source spectra. If this is unknown, can use

$$w(E, \varphi) = \frac{\theta f_S(\varphi|E)}{\theta f_S(\varphi|E) + (1-\theta) f_B(\varphi|E)}$$

Weight function

The weight function depends on θ, the proportion of photons from the source. This may be known from previous studies. It might also be estimated from the log likelihood function under the null hypothesis (source is not periodic)

$$\ell(\theta) = N \log \mu + \sum_{j=1}^{N} \log c(t_j) + \sum_{j=1}^{N} \log[(1-\theta) f_B(z_j) + \theta f_S(z_j)] - \mu \int_0^T c(t)\,dt.$$

For a weak source (small θ)

$$w \approx \frac{f_S(E) f_S(\varphi|E)}{f_B(E) f_B(\varphi|E)}$$

Detection sensitivity: power of the test

Let the pulse shape of the source be

$$\gamma(t) = \sum_{n \neq 0} \gamma_n e^{2\pi i n t}$$

Then,

Signal to Noise Ratio

$$\frac{E_K(Q_T) - E_H(Q_T)}{\sigma_H} \simeq \theta^2 T \mu_0 \mathcal{E}(w) \frac{|\gamma_n|^2 |\alpha_n|^2}{[\sum_{n \neq 0} |\alpha_n|^4]^{1/2}}$$

where H denotes the null hypothesis and K the alternative hypothesis and for any weight function $w(z)$

$$\mathcal{E}(w) = \frac{[E(W|\text{Source})]^2}{E(W^2)}$$

Power

$$\frac{E_K(Q_T) - E_H(Q_T)}{\sigma_H} \simeq \theta^2 T \mu_0 \mathcal{E}(w) \frac{\sum_{n\neq 0} |\gamma_n|^2 |\alpha_n|^2}{[\sum_{n\neq 0} |\alpha_n|^4]^{1/2}}$$

$\theta^2 T$: threshold for detecting weak signal is $\theta \sim T^{-1/2}$

$\sum_{n\neq 0} |\gamma_n|^2 |\alpha_n|^2$: corr of actual lightcurve with template

$\mathcal{E}(w) = \frac{[E(WW_{\text{opt}})]^2}{E(W^2)}$: correlation of weight function with optimal weight function

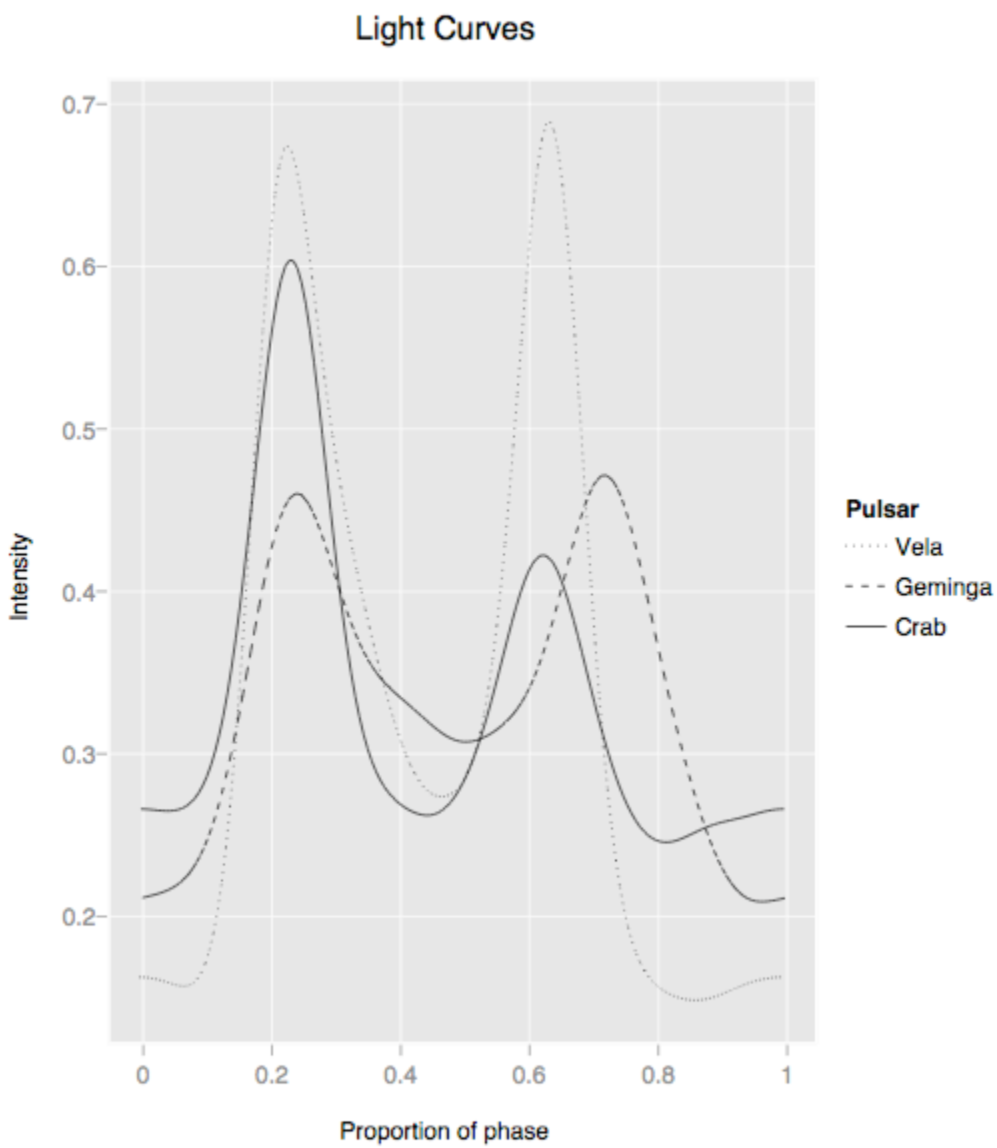
Power: effect of frequency misspecification

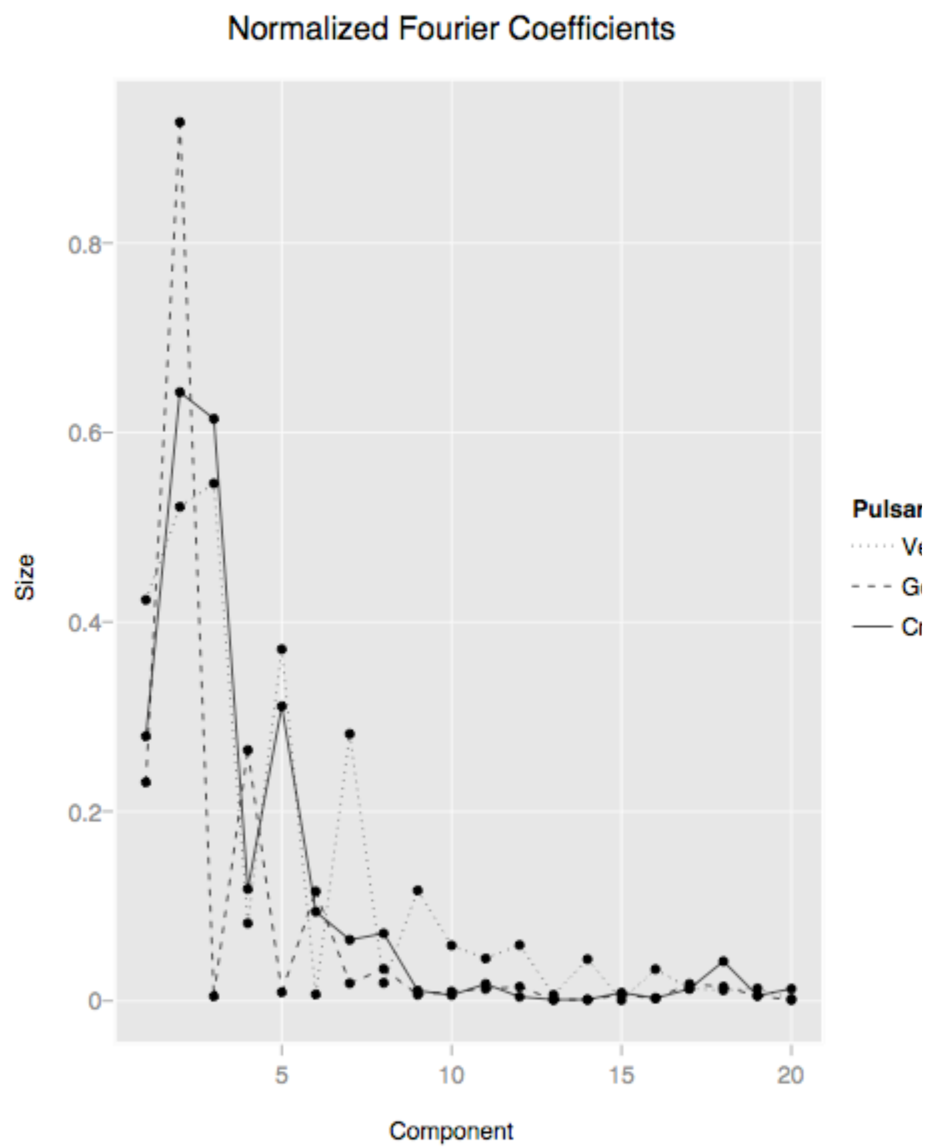
$$f = f_0 + \Delta/T \text{ and } \dot{f} = \dot{f}_0 + \Delta/T^2, \Delta < 1$$

$$\sum_{n \neq 0} |\gamma_n|^2 |\alpha_n|^2 (1 + O((n\Delta)^2))$$

High accuracy is required to gain power from higher harmonics

Example: template





Example: template

Table 1: Relative efficiencies for truncated equal weights

number of coefficients	1	2	3	4	5	6	7	8	9
Crab	28	65	89	83	88	84	80	78	74
Geminga	23	82	67	71	66	63	59	57	54
Vela	42	67	86	79	87	80	84	79	79

Table 2: Relative efficiencies obtained from using the first five and first ten average coefficients as the template.

number of terms	5	10
Crab	96	97
Geminga	85	85
Vela	89	93

Example: weight function

Consider a source which emits photons at rate α and a background whose rate is ρ per unit area and suppose that photons are collected in a disc of radius R (rather than a spherical cap, for simplicity). Then

$$\begin{aligned}
\mu &= \pi R^2 \rho + \alpha \\
\theta &= \frac{\alpha}{\pi R^2 \rho + \alpha} \\
f_B(\varphi|E) &= \frac{2\varphi}{R^2}, \quad 0 \leq \varphi \leq R
\end{aligned}$$

The optimal weight function is then

$$w_{\mathrm{opt}}(E,\varphi) = \frac{f_S(E) f_S(\varphi|E)}{f_S(E) f_S(\varphi|E) + \beta\varphi f_B(E)}$$

where $\beta = 2\pi\rho/\alpha$

Example: weight function

If the psf is bivariate circular Gaussian with standard deviation $\sigma(E)$, then φ, the distance to the origin, has the probability density function

$$f_S(\varphi|E) = \frac{\varphi}{\sigma(E)} \exp(-\frac{\varphi^2}{2\sigma^2(E)})$$

(This assumes that $\sigma(E) \ll R$, otherwise the density truncated at R has to be normalized to have unit area.) Then the optimal weight function is

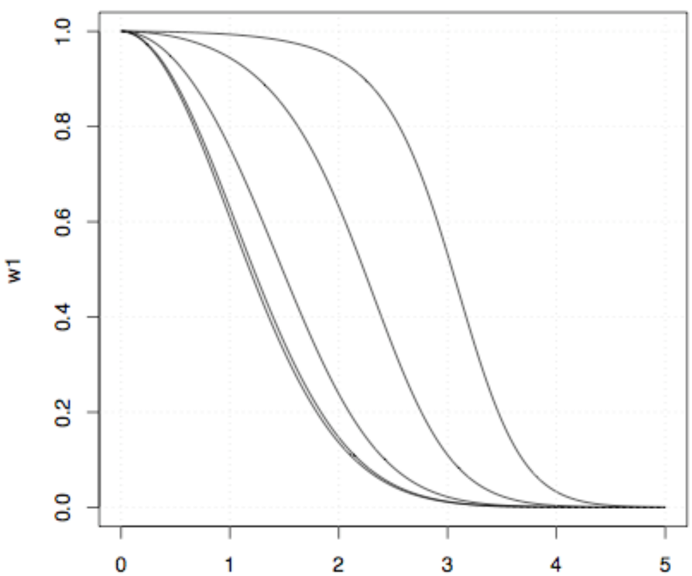
$$w_{\text{opt}}(E, \varphi) = \frac{f_S(E)}{f_S(E) + \beta\sigma(E)\exp(\varphi^2/2\sigma^2(E))f_B(E)}$$

Example: weight function

If photons are not differentially weighted according to the ratio of the energy spectra, one has the weight function

$$w(E, \varphi) = \frac{1}{1 + \beta\sigma(E)\exp(\varphi^2/2\sigma^2(E))}$$

The decay of the weight function depends on the parameter $\xi = 2\pi\rho\sigma(E)/\alpha$.



$\xi = 100, 10, 1, 0.1, 0.01$

Difficulties

Frequency unknown
Spin down
Large search space
Glitches
Celestial foreground
Barycentric time correction
Pulse profile unknown

Computational demands for a blind search are very substantial. A heroic search using a 512 processor supercomputer did not find any previously unknown gamma-ray pulsars in EGRET data. (Chandler et al, 2001).

Search Space

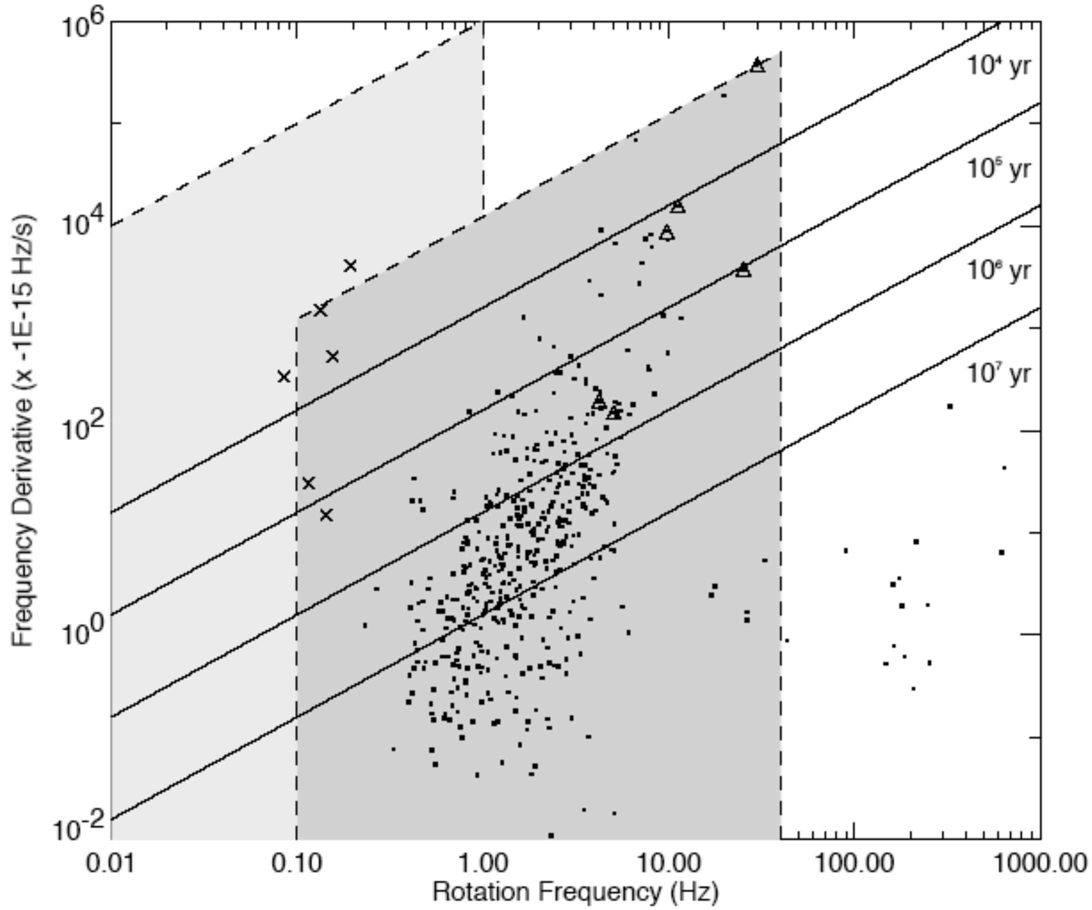


FIG. 1.—The f-$\dot{f}$ phase space relevant to our search. The 472 pulsars in the Princeton catalog (Taylor et al. 1993) that have positive period derivatives are denoted by filled circles. The known EGRET pulsars are indicated by triangles. They are, in order of decreasing rotation frequency, Crab, PSR 1951+32, Vela, PSR 1706−44, Geminga, and PSR 1055−52.

From Chandler et al (2001)

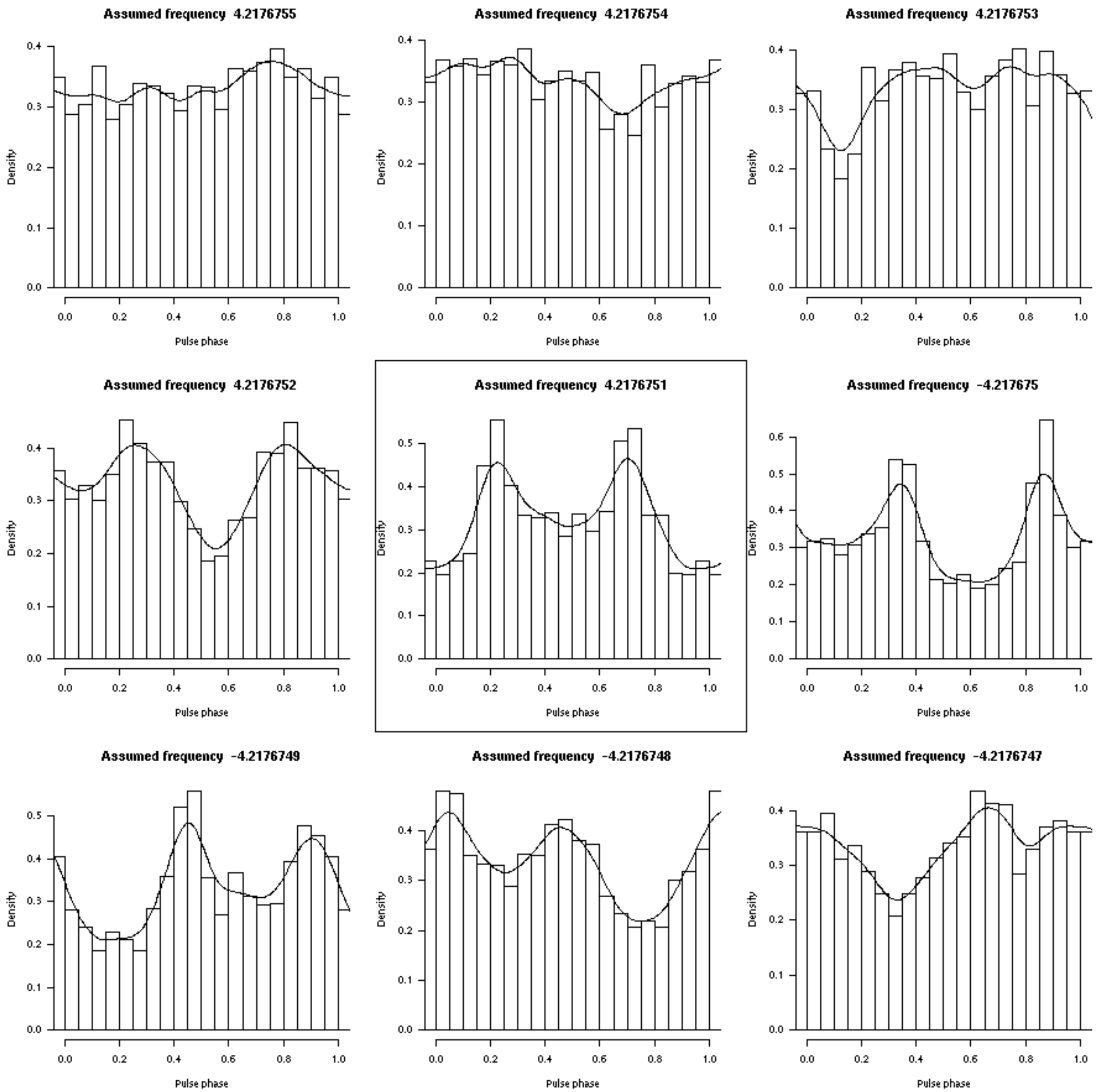
Search Space

Consider no drift. Good frequency resolution depends on matching phase of photons at beginning and end of the record. If true frequency is f_0, the number of cycles in time T is T/f_0, so if the hypothesized frequency is $f = f_0 + \delta f$, δf should be $o(T^{-1})$ in order for a photon at the end of the record to be in phase with one at the beginning. The phase error at the end of the record is $T\delta f$.

10 days = 864,000 sec,- $\delta f = T^{-1.}$ If a 40 Hz range has to be searched, a minimum of 40 x 864000 = 34,560,000 possible frequencies must be examined.

Similarly, drift must be resolved within $o(T^{-2})$. To search the interval of possible frequency derivatives at this resolution, about 400-500 values must be examined.

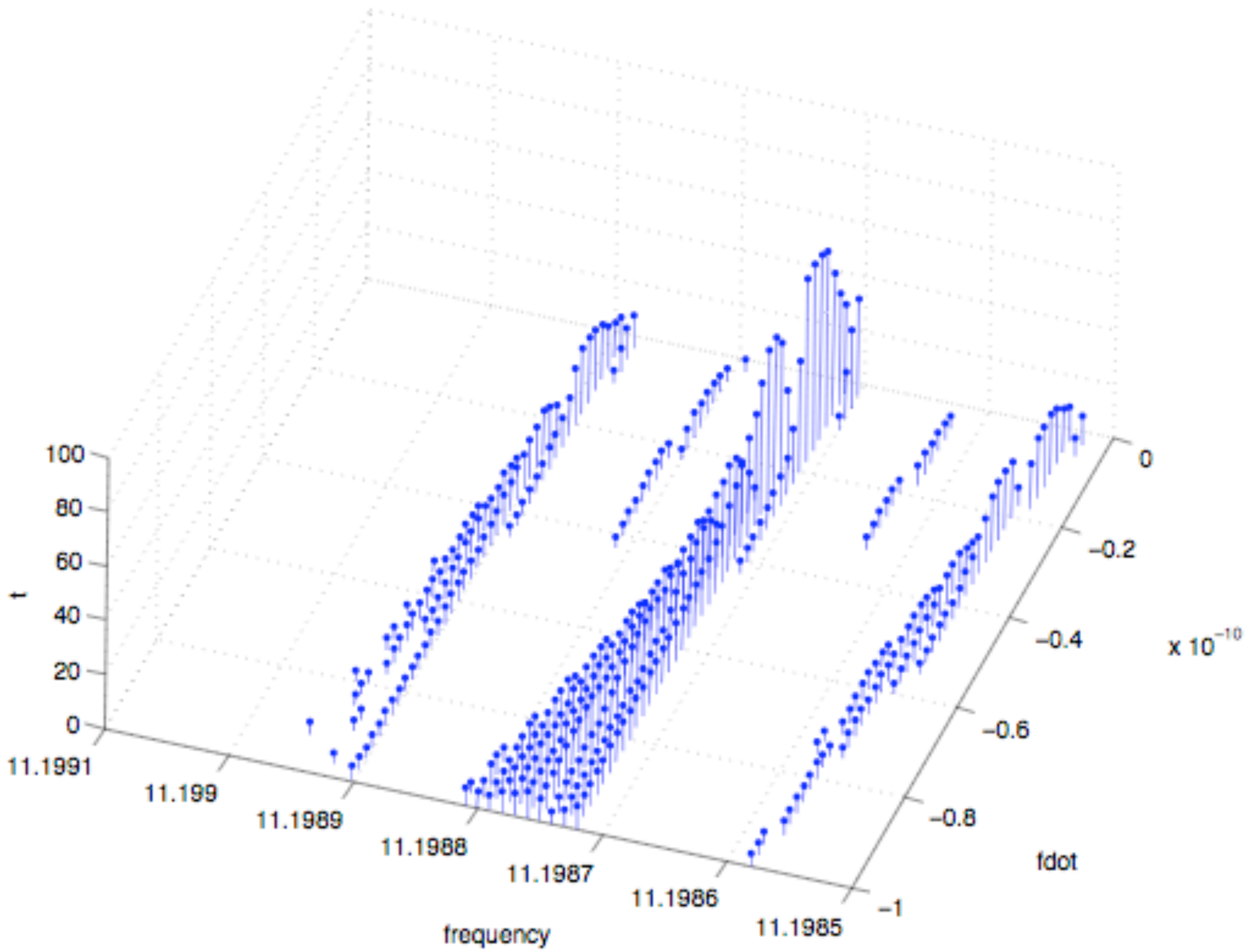
Consequence is that a test statistic must be evaluated $\sim 10^9$ values of frequency and its derivative.



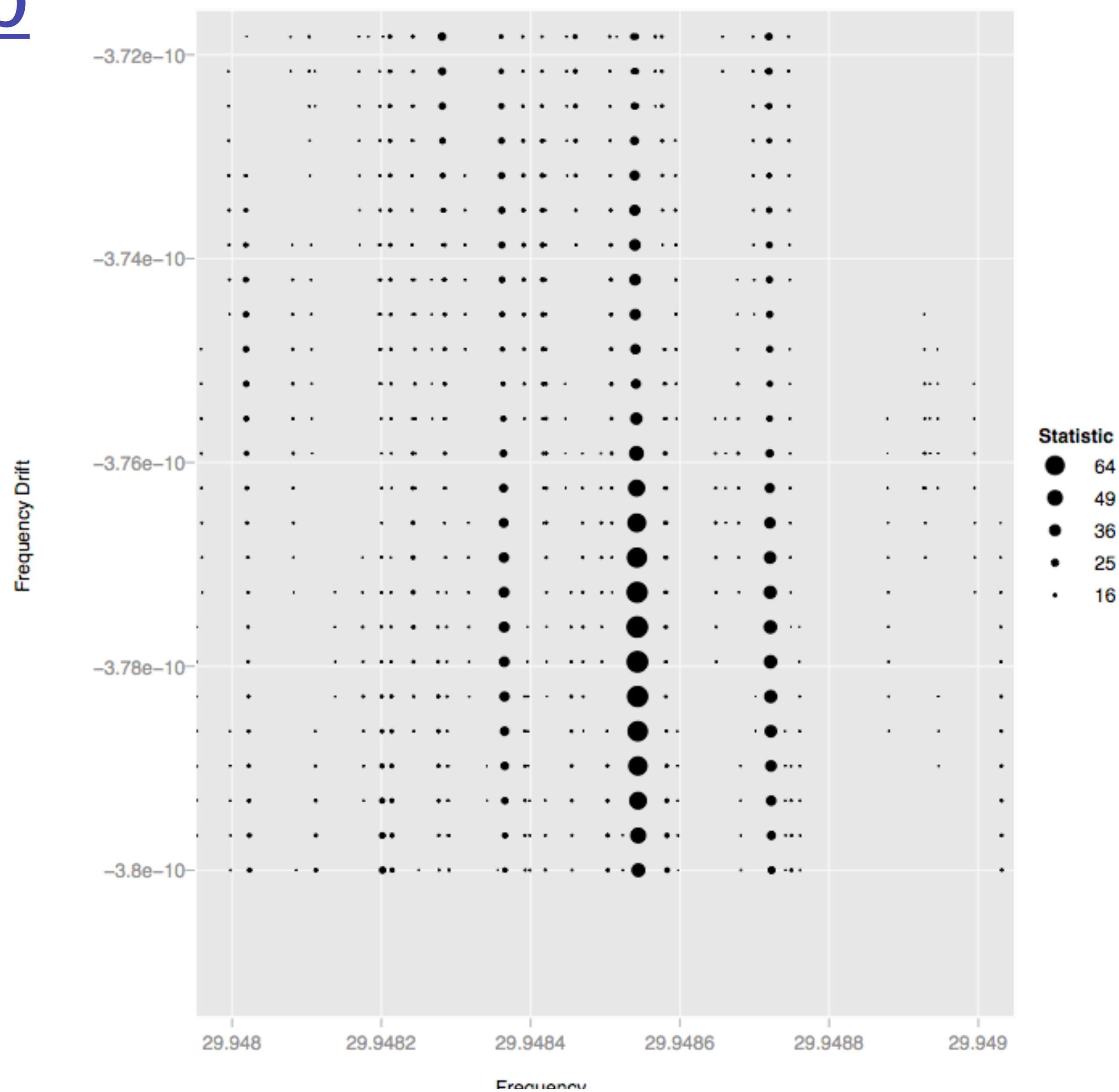
Histograms and density estimates for phased and folded arrival times from Geminga

t* = t mod P (P = 1/f) for increments in f of size 10^{-7} (.12/T)

Example: Vela



Crab



-3.72e-10
-3.74e-10
-3.76e-10
-3.78e-10
-3.8e-10
Frequency Drift
29.948
29.9482
29.9484
29.9486
29.9488
29.949
Frequency
Statistic
64
49
36
25
16

Power vs Computational Cost

$\text{Power} \propto \theta^2 T$

f resolution $\propto T^{-1}$

$\dot{f}$ resolution $\propto T^{-2}$

Calculation of statistic for a single $(f, \dot{f}) \propto T$

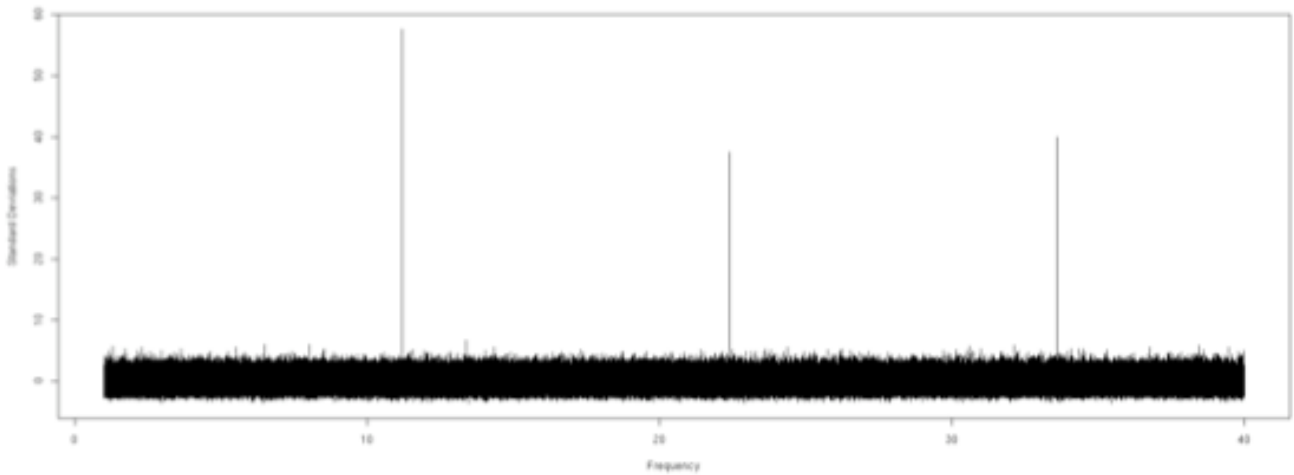
$\text{FLOPS} \propto T^4$

Partition T into B blocks of length σ. Compute statistic in each block and average.

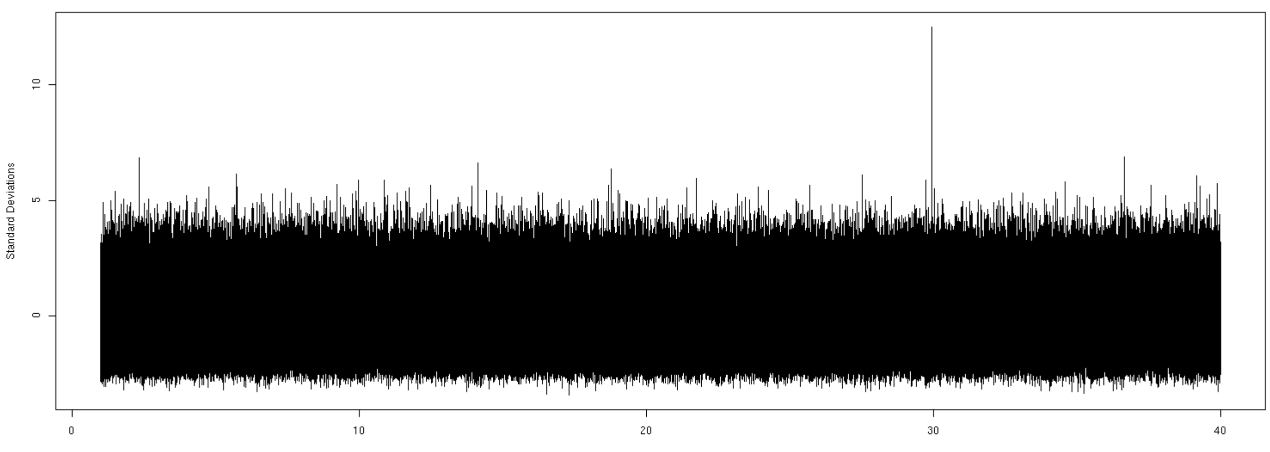
$\text{Power} \propto \frac{\theta^2 T}{\sqrt{B}} = \theta^2 \sigma^{1/2} T^{1/2}$

$\text{FLOPS} \propto \sigma T$

Blocking Vela and Crab



Vela: 318 blocks



Crab: 25 blocks

Integration versus discretization

Rather than fine discretization of frequency, consider integrating the test statistic over a frequency band using a symmetric probability density *g(f)*.

$$
\begin{aligned}
\bar{Q_n} &= \int |A_n(f)|^2 g(f)\,df \\
&= \sum_j \sum_k w_j w_k \hat{g}(2\pi n(t_j - t_k))
\end{aligned}
$$

Decreasing function of $|t_j$ -

where

$$\hat{g}(t) = \int_{-\infty}^{\infty} e^{itf} g(f)\,df$$

$$\bar{Q}_n = \sum_j \sum_k w_j w_k \hat{g}(2\pi n(t_j - t_k))$$

Requires a number of operations quadratic in the number of photons. However the quadratic form can be diagonalized in an eigenfunction expansion, resulting in a number of operations linear in the number of photons.

$$\bar{Q}_n = \sum_k \mu_{kn} \left| \int_0^T e^{2\pi i n t f_0} \psi_{kn}(t) dW(t) \right|^2$$

(In the case that *g*() is uniform, the eigenfunctions are the prolate spheroidal wave functions.) Then

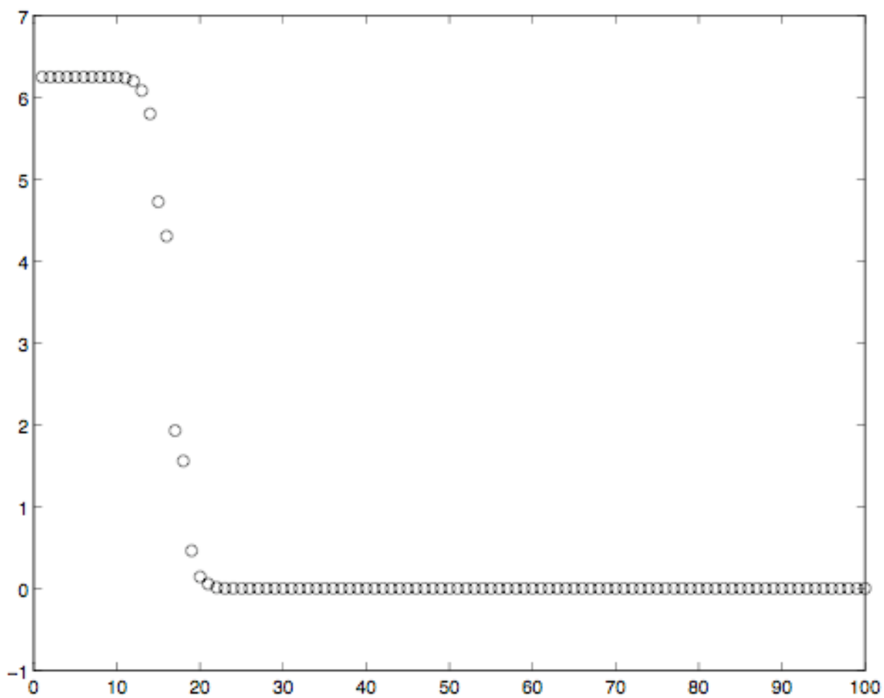
$$E\bar{Q}_T = \mu^2 \theta^2 T \sum_n |\alpha_n|^2 |\gamma_n|^2 \int \left| \frac{\sin(n\pi u)}{n\pi u} \right|^2 g_0(u) du$$

$$g(f) = T g_0(T(f - f_0))$$

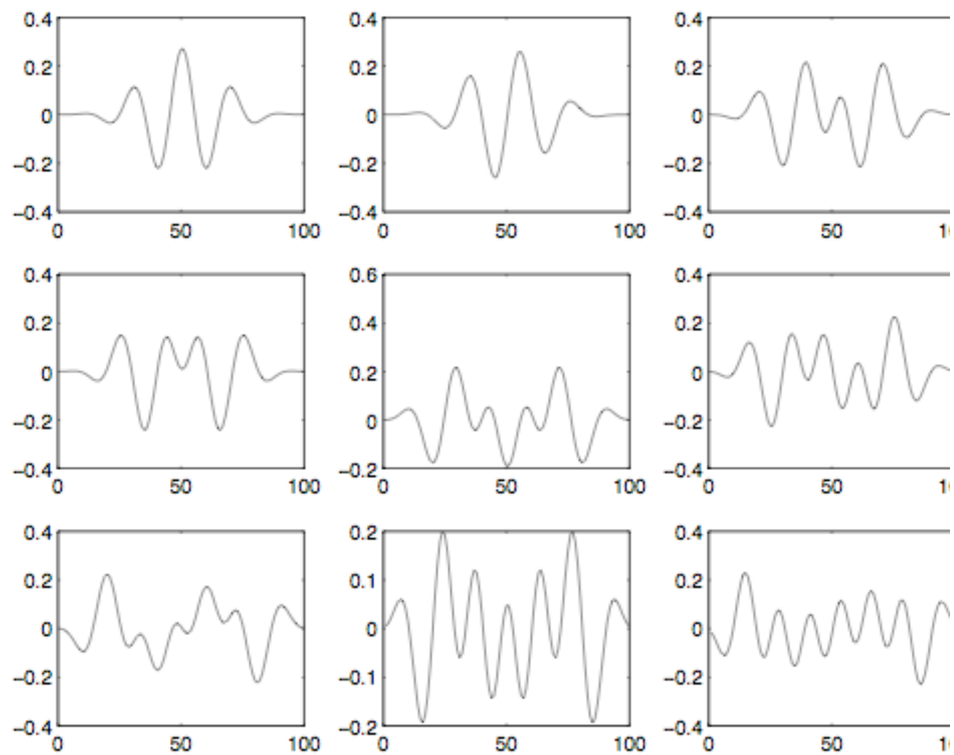
Power is still lost in high frequencies unless the support of *g* is small.

This procedure can be extended to integrate over tiles in the $(f, \dot{f})$ plane when

$$\phi(t) = ft + \frac{1}{2}\dot{f}t^2$$



7
6
5
4
3
2
1
0
−1
0
10
20
30
40
50
60
70
80
90
100



0.4
0.2
0
−0.2
−0.4
0.6
0.1
0
50
100

Assessing significance

At a single frequency, significance can be assessed easily through simulation. In a broadband blind search this is not feasible and furthermore one may feel nervous in using the traditional chi-square approximations in the extreme tail (it can be shown that the limiting null distribution of the integrated test statistic is that of a weighted sum of chi-square random variables). We are thus investigating the use of classical extreme value theory in conjunction with affordable simulation.

Gumbel Approximation

Consider $M_n = \max\{T_1, T_2, \ldots, T_n\}$ where the T_i are iid random variables. Classical theory gives that if M_n has a limiting distribution it is of one of three types. In the case of mixtures of chi-square random variables, the limiting distribution will be of Gumbel type.

$$F^n(a_n t + b_n) \to \exp[-\exp(-x)]$$

$$P(T > t + s | T > t) \approx e^{-\lambda s}$$

We thus approximate tail probabilities by

$$P(T > t + s) \approx P_n(T > t) e^{-s/\alpha_n}$$

where P_n is the empirical measure and α_n is the mean of the corresponding empirical conditional distribution.

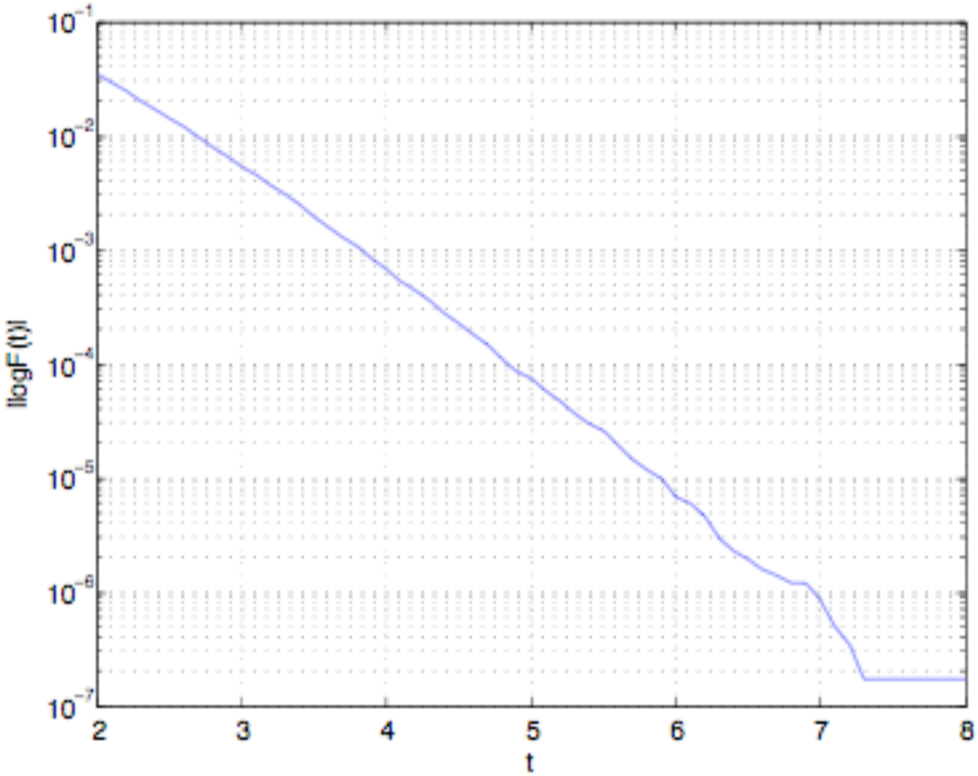
Tail Approximations

Table 1: Conditional means and tail probability approximations for various cuttoff values t.

t	α	$P(T > 10)$	$P(T > 15)$
3	.4852	3.0×10^{-9}	9.9×10^{-14}
4	.4548	1.3×10^{-9}	2.1×10^{-14}
5	.4438	9.4×10^{-10}	1.2×10^{-14}

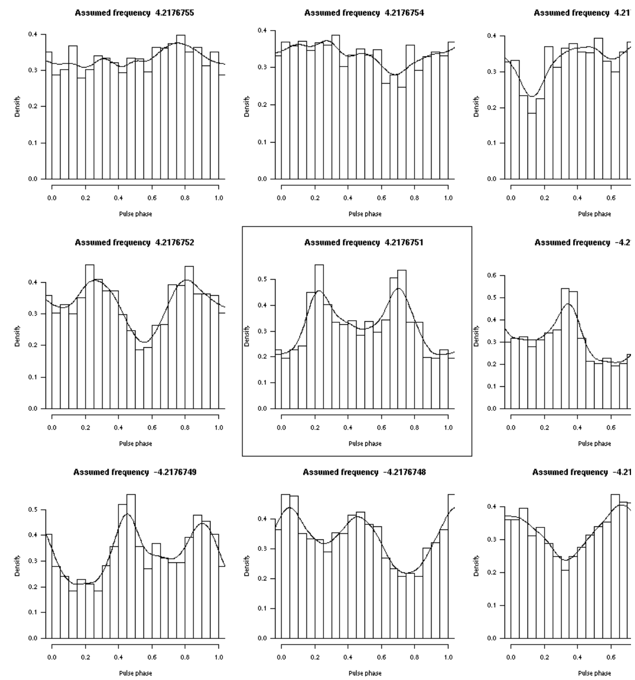
According to this approximation, in order for a Bonferonni corrected p-value to be less than 0.01, a test statistic of about 11 standard deviations or more would be required.

log[- log F(t)] versus t

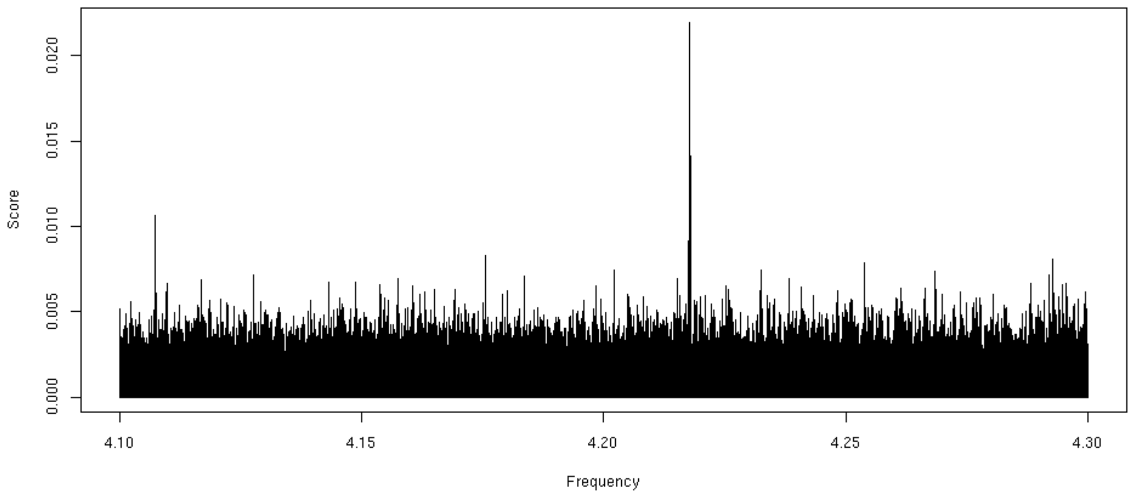


Thank you

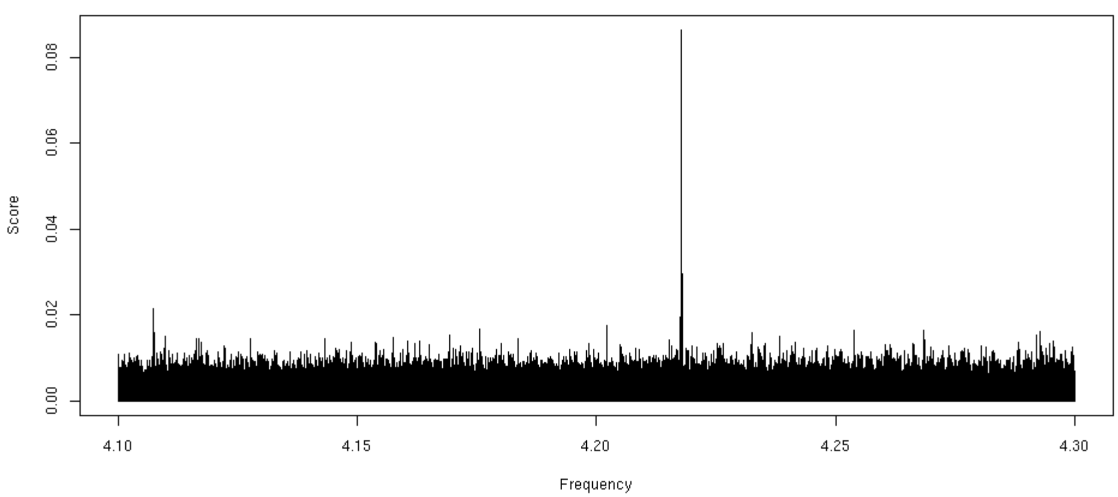
Example: Geminga

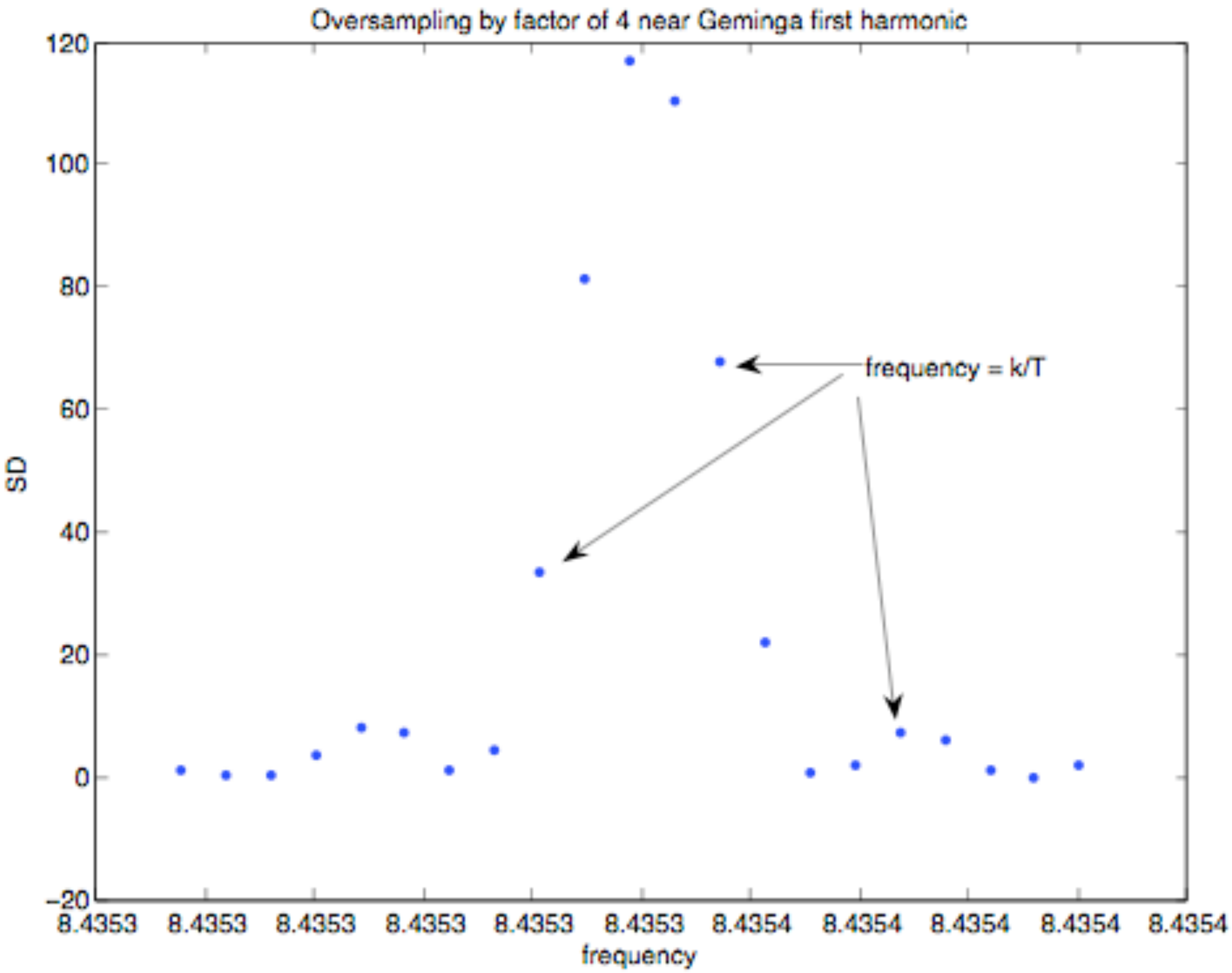


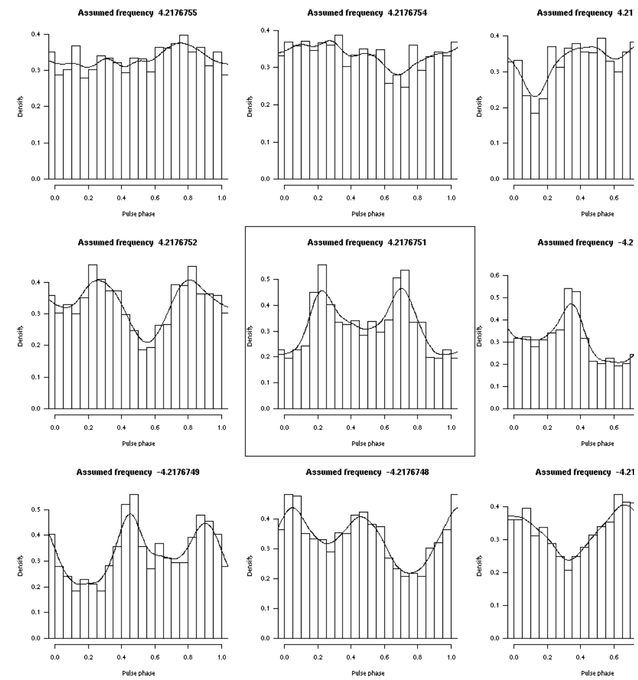
Main frequency only, N=1



Main frequency plus equally weighted first harmonic




Oversampling by factor of 4 near Geminga first harmonic
SD
frequency
frequency = k/T
120
100
80
60
40
20
0
−20
8.4353
8.4354


Assumed frequency 4.2176755
Assumed frequency 4.2176754
Assumed frequency 4.21
Assumed frequency 4.2176752
Assumed frequency 4.2176751
Assumed frequency -4.2
Assumed frequency -4.2176749
Assumed frequency -4.2176748
Assumed frequency -4.21
Density
Pulse phase