Resolving multiple SMBH binaries with pulsar timing arrays.

Stas Babak, Antoine Petiteau, Alberto Sesana

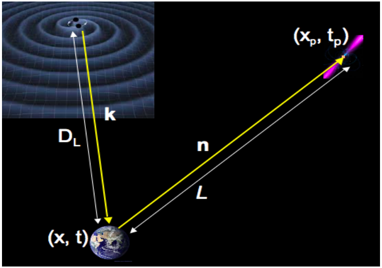
Albert Einstein Institute (Golm)

Outline

Response to GW

The response to GW is given as

$$\delta\tau_{GW} = r(t) = \int_0^t \frac{\delta\nu}{\nu}(t')dt'; \qquad \frac{\delta\nu}{\nu} = \frac{1}{2}\frac{\hat{n}^i\hat{n}^j}{1+\hat{n}.\hat{k}}\Delta h_{ij}$$



$\Delta h_{ij} = h_{ij}(t_p = t - L(1+\hat{n}.\hat{k})) - h_{ij}(t)$
Since the pulsars are not correlated (t_p, the emission time of the pulse detected at the time t on the Earth, is different for all pulsars) the "pulsar" terms do not add up coherently.

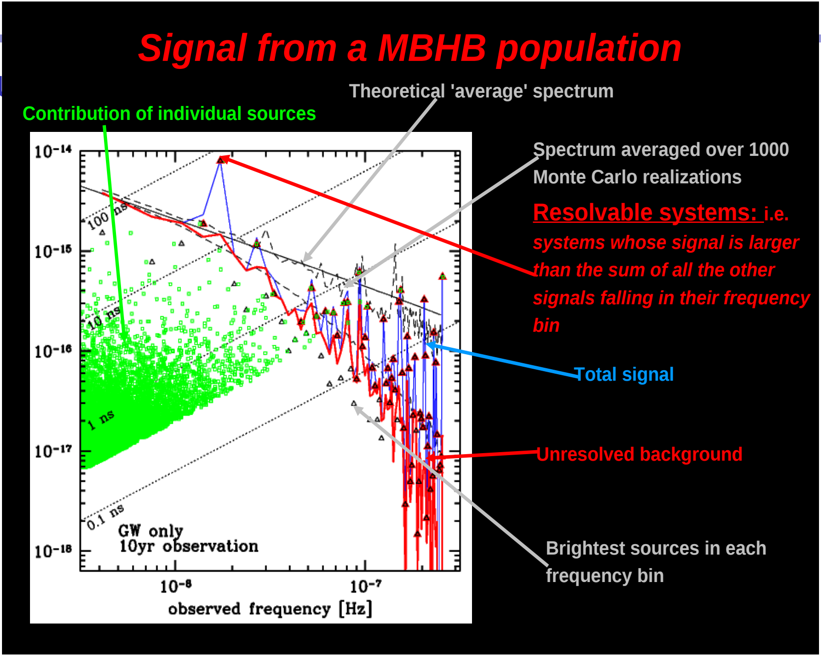
Sources

- Stochastic gravitational waves (cosmological origin, cosmic strings, ...)

Sources

- Stochastic gravitational waves (cosmological origin, cosmic strings, ...)
- SuperMassive Black Hole binaries

Sou


Signal from a MBHB population
Theoretical 'average' spectrum
Contribution of individual sources
Spectrum averaged over 1000 Monte Carlo realizations
Resolvable systems: i.e. systems whose signal is larger than the sum of all the other signals falling in their frequency bin
Total signal
Unresolved background
Brightest sources in each frequency bin
100 ns
10 ns
1 ns
0.1 ns
GW only
10yr observation
10^{-14}
10^{-15}
10^{-16}
10^{-17}
10^{-18}
10^{-8}
10^{-7}
observed frequency [Hz]

Assumptions

- We assume that the data (observations) are evenly sampled

Assumptions

- We assume that the data (observations) are evenly sampled
- We neglect all stochastic backgrounds and concentrate on the resolving individual sources (SMBH binaries)

Assumptions

- We assume that the data (observations) are evenly sampled
- We neglect all stochastic backgrounds and concentrate on the resolving individual sources (SMBH binaries)
- We assume that the frequency of GW doesn't change appreciably over the observation time ($\sim 10 years$).

Assumptions

- We assume that the data (observations) are evenly sampled
- We neglect all stochastic backgrounds and concentrate on the resolving individual sources (SMBH binaries)
- We assume that the frequency of GW doesn't change appreciably over the observation time ($\sim 10 years$).
- Source location and location of pulsars are drawn random on the sky

Assumptions

- We assume that the data (observations) are evenly sampled
- We neglect all stochastic backgrounds and concentrate on the resolving individual sources (SMBH binaries)
- We assume that the frequency of GW doesn't change appreciably over the observation time ($\sim 10 years$).
- Source location and location of pulsars are drawn random on the sky
- Assume in our simulations equal (50 ns) white noise in all pulsars

Assumptions

- We assume that the data (observations) are evenly sampled
- We neglect all stochastic backgrounds and concentrate on the resolving individual sources (SMBH binaries)
- We assume that the frequency of GW doesn't change appreciably over the observation time ($\sim 10 years$).
- Source location and location of pulsars are drawn random on the sky
- Assume in our simulations equal (50 ns) white noise in all pulsars
- We consider three data sets: (i) 5 sources with SNR (50-70), (ii) 4 sources with SNR (10-50), (iii) 8 sources with SNR (10-30)

Assumptions

- We assume that the data (observations) are evenly sampled
- We neglect all stochastic backgrounds and concentrate on the resolving individual sources (SMBH binaries)
- We assume that the frequency of GW doesn't change appreciably over the observation time ($\sim 10 years$).
- Source location and location of pulsars are drawn random on the sky
- Assume in our simulations equal (50 ns) white noise in all pulsars
- We consider three data sets: (i) 5 sources with SNR (50-70), (ii) 4 sources with SNR (10-50), (iii) 8 sources with SNR (10-30)
- We build detection statistic based on the earth term only! We treat the pulsar terms in the data as sources of noise

Detection statistic

Here we **use the “earth” term only**, so $r_\alpha \to r_\alpha^E$ The residuals in t.o.a caused by a single GW signal could be presented as

$$r_\alpha(t) = \sum_{j=1}^{4} a_{(j)} h_{(j)}^\alpha$$

$$a_{(j)} = a_{(j)}(\iota, \psi, \phi_0, \mathcal{A}), \quad h_{(j)}^\alpha = h_{(j)}^\alpha(t, f, \hat{n}_\alpha, \theta, \phi).$$

Consider an observed data set of residuals in t.o.a. x_α, then the log-likelihood that this data set contains a GW signal $r_\alpha(t; \vec{\lambda})$ (here $\vec{\lambda}$ are parameters of GW source) is

$$\log \Lambda_\alpha \sim (x_\alpha || r_\alpha) - \frac{1}{2}(r_\alpha || r_\alpha)$$

$$(x||h) \equiv \frac{2}{T_o}\int_0^{T_o} x(t)h(t)dt$$

Detection statistic

Here we **use the "earth" term only**, so $r_\alpha \to r_\alpha^E$ The residuals in t.o.a caused by a single GW signal could be presented as

$$r_\alpha(t) = \sum_{j=1}^{4} a_{(j)} h_{(j)}^{\alpha}$$

$$a_{(j)} = a_{(j)}(\iota, \psi, \phi_0, \mathcal{A}), \quad h_{(j)}^{\alpha} = h_{(j)}^{\alpha}(t, f, \hat{n}_\alpha, \theta, \phi).$$

Consider an observed data set of residuals in t.o.a. x_α, then the log-likelihood that this data set contains a GW signal $r_\alpha(t; \vec{\lambda})$ (here $\vec{\lambda}$ are parameters of GW source) is

$$\log \Lambda_\alpha \sim (x_\alpha || r_\alpha) - \frac{1}{2}(r_\alpha || r_\alpha)$$

$$(x||h) \equiv \frac{2}{T_o} \int_0^{T_o} x(t) h(t) dt$$

The total log-likelihood is a sum over each pulsar data, we also take into account the explicit form of r^α to bring into the form similar to F-statistic (used in detecting monochromatic GW signals from pulsars).

$$\log \Lambda = \sum_{\alpha=1}^{P} \log \Lambda_\alpha \sim \sum_{j=1}^{4} a_{(j)} X_j - \frac{1}{2} \sum_{j=1}^{4} \sum_{k=1}^{4} a_{(j)} a_{(k)} M_{jk}$$

$$X_j \equiv \sum_{\alpha=1}^{P} (x_\alpha || h^\alpha_{(j)}), \quad M_{ik} \equiv \sum_{\alpha=1}^{P} (h^\alpha_{(j)} || h^\alpha_{(k)}).$$

Here $a_{(j)}, X_j$ are $4N$-dimensional vectors, M_{jkj} is $4N \times 4N$ matrix, N - is a number of GW sources (unknown). This allows us to maximize over $a_{(j)}$ analytically

Parameters count: GW signal: $\mathcal{A}, \iota, \psi, \phi_0$, f, θ, ϕ: 6 parameters (+frequency).

- The residual data from each pulsar at a given frequency provides us with 2 measures: amplitude and phase $\rightarrow$ need $3 \times N$ pulsars to estimate all six parameters of N GW sources

Parameters count: GW signal: $\mathcal{A}, \iota, \psi, \phi_0$, f, θ, ϕ: 6 parameters (+frequency).

- The residual data from each pulsar at a given frequency provides us with 2 measures: amplitude and phase $\rightarrow$ need $3 \times N$ pulsars to estimate all six parameters of N GW sources
- Adding frequency as a search parameter doesn't need extra pulsars: Fourier transform of the data $\rightarrow$ gives two parameters (amplitude and phase) at each Fourier frequency bin $\rightarrow$ the signals could be treated independently if freqs. of GW signals separated by more than few Fourier bins.

Parameters count: GW signal: $\mathcal{A}, \iota, \psi, \phi_0$, f, θ, ϕ: 6 parameters (+frequency).

- The residual data from each pulsar at a given frequency provides us with 2 measures: amplitude and phase $\rightarrow$ need $3 \times N$ pulsars to estimate all six parameters of N GW sources
- Adding frequency as a search parameter doesn't need extra pulsars: Fourier transform of the data $\rightarrow$ gives two parameters (amplitude and phase) at each Fourier frequency bin $\rightarrow$ the signals could be treated independently if freqs. of GW signals separated by more than few Fourier bins.
- The limit of frequency resolution of different GW signals is $\sim 2/3 \Delta F$

Search with Genetic Algorithm

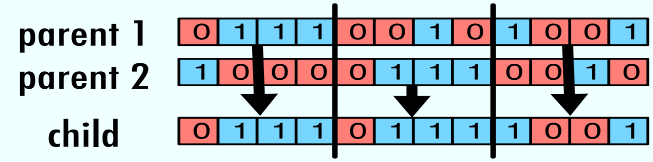
- Genetic Algorithm is quite common optimization technique based on the Darwin's natural selection principle
- We have colony of organisms characterized by "fitness" - likelihood
- strong organisms (high likelihood) survive during the evolution and give life to a new generation
- Each organism have set of genes - parameters (number of GW sources, frequencies, sky location)

Genetic algorithm		GW search
organism	$\Longleftrightarrow$	template
gene (of an organism)	$\Longleftrightarrow$	parameter (of a template)
allele (of a gene)	$\Longleftrightarrow$	bits (of the value of the parameter)
quality Q	$\Longleftrightarrow$	Maximized Likelihood or $\mathcal{A}$-statistic
colony of organisms	$\Longleftrightarrow$	evolving group of templates
n-th generation	$\Longleftrightarrow$	the state of colony at n-th step of evolution
(selection + breeding) + mutation	$\Longleftrightarrow$	way of exploring the parameter space

Genetic Algorithm

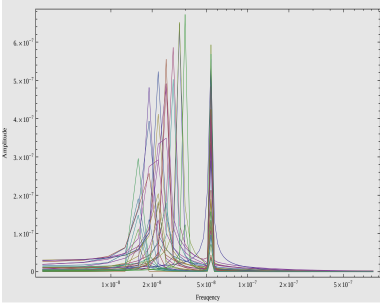
There are three basic operations which define evolution of a colony

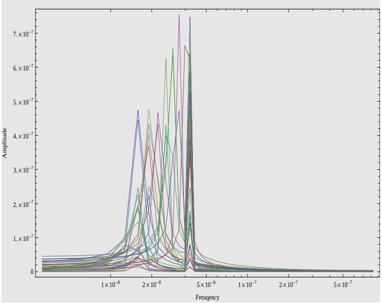
- **Selection**: we select two (or three) parents from a current generation for breeding: it is based on the quality of an organism - the higher likelihood the most likely the organism will be chosen for breeding
- **Breeding**: the rule which we apply to produce a child out of parents (many ways to do that)
- **Mutation**: we randomly change some or all prameters of a new generation with some probability



Pulsar-terms

(i) Pulsar terms fall at different (lower) frequencies and do not add up coherently (ii) we treat them as sources of noise (non-Gaussian features) (iii) But(!) $r^P_\alpha \sim \omega_\alpha^{-1/3}$: usually higher amplitude than "earth term".





We construct the following correlations:

$$SNe[\alpha; i] = (r_i^\alpha | \sum_i^{N_s} r_i^\alpha)$$

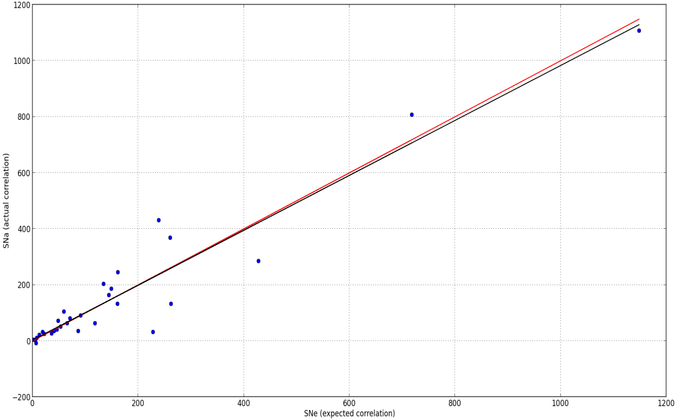
- correlation (expected) between expected contribution from the GW source "i" to the total (expected) signal at pulsar α, **using earth term only**; and

$$SNa[\alpha; i] = (r_i^\alpha | d^\alpha)$$

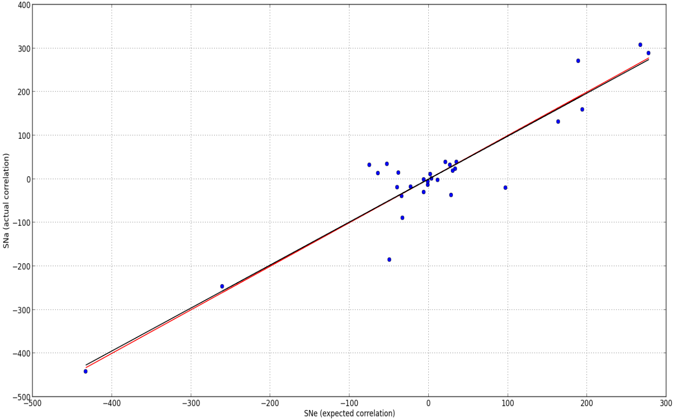
- actual correlation, between expected individual contribution from each GW source and the data $d = r^E + r^P + n$. We do it for each GW source candidate found by the search algorithm.

Comparing SNe vs SNa for earth term and pulsar term generated GW candidate

Earth term. Can also measure correlation coefficient r^2.



Comparing SNe vs SNa for earth term and pulsar term generated GW candidate



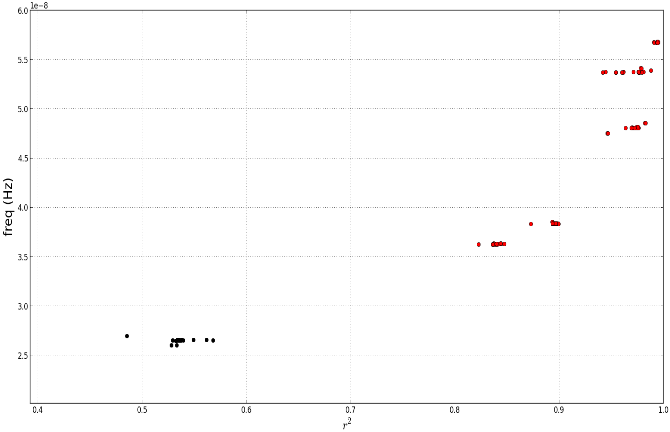
Pulsar term

Using high/band pass filter

- The high frequency sources are usually free of the "pulsar term corruption", but weaker $r \sim \omega(t)^{-1/3}$
- We apply series of high pass filters to recover the high frequency sources (20, 40, 60, 80, 100, 120) nHz and analyze each processed data set separately
- The filter is very broad (transition band $\Delta f \sim 40 - 60$nHz).
- On each band limited data set we search for *Ns* (number of GW sources) and for each source search for θ, ϕ, f (marginalizing likelihood over other parameters analytically.

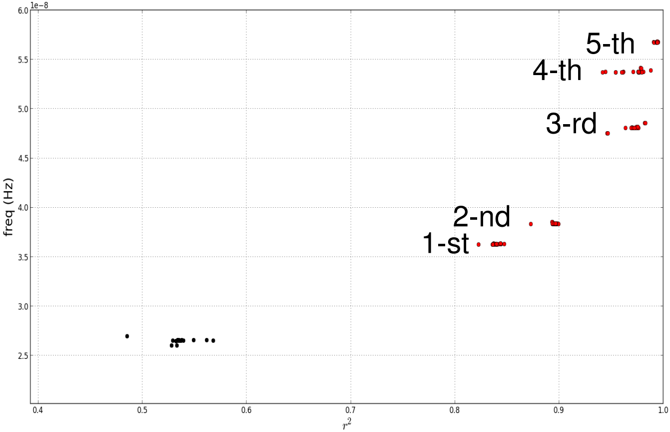
Dataset 1

The sources were strong: we didn't need to use filters.



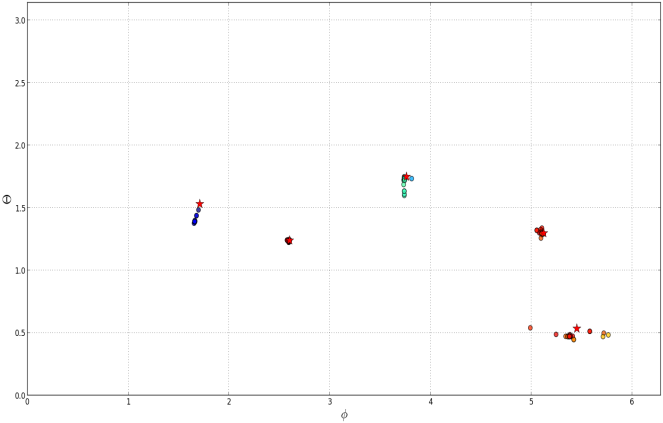
Dataset 1

The sources were strong: we didn't need to use filters.



Dataset 1

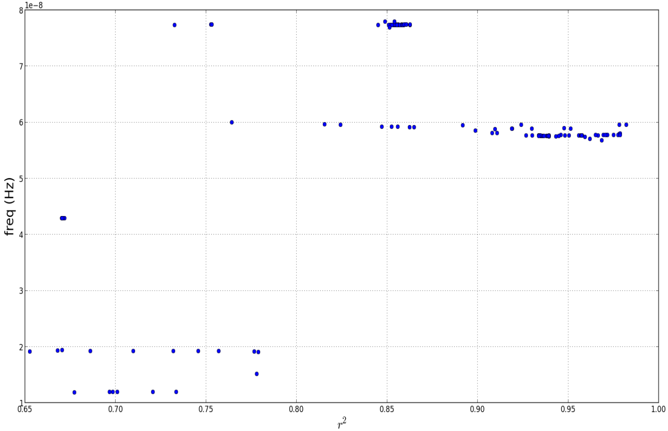
The sources were strong: we didn't need to use filters



r^2

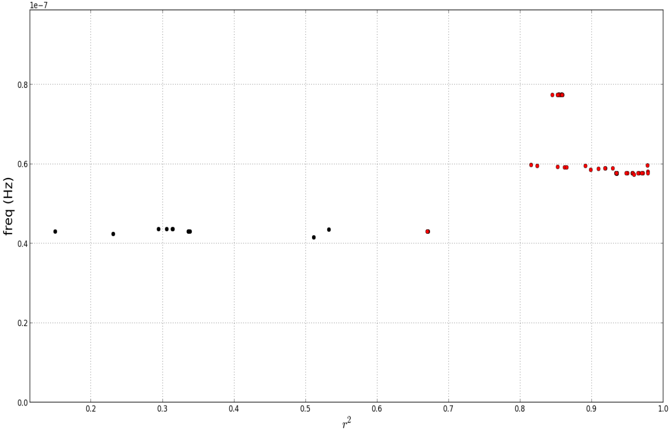
Dataset 2

The sources were weak: we did use filters.



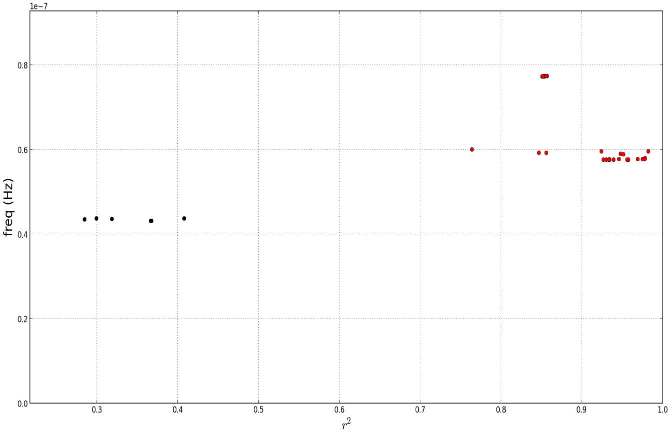
Dataset 2

60 nHz high pass filter



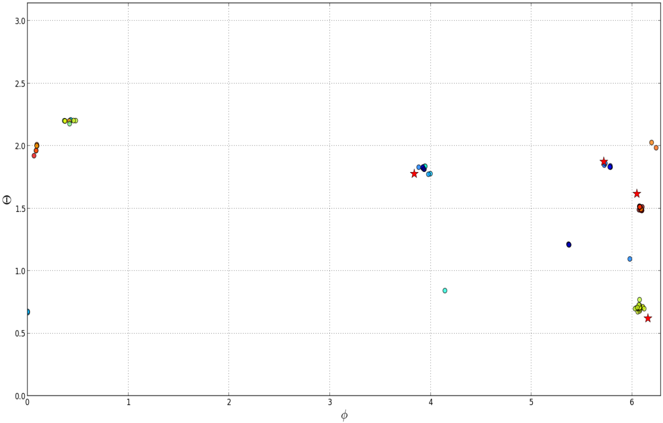
Dataset 2

80 nHz high pass filter



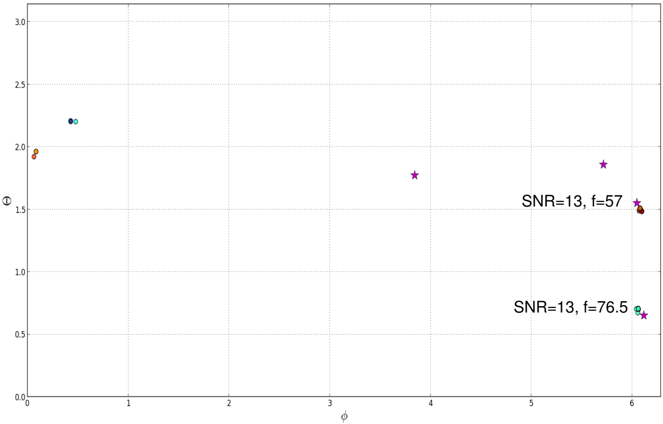
Dataset 2

The sources were weak: we did use filters.



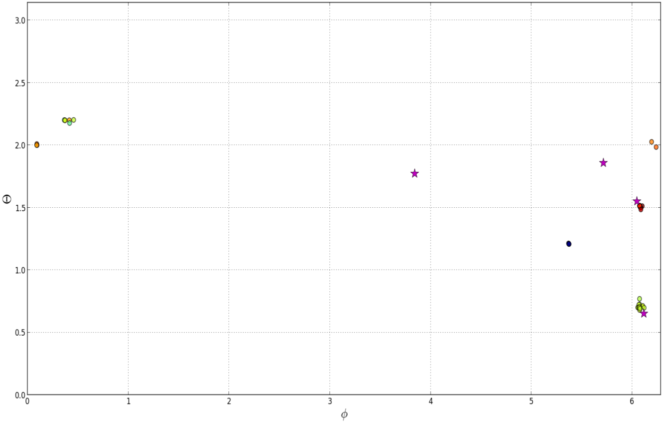
Dataset 2

80 nHz high pass filter



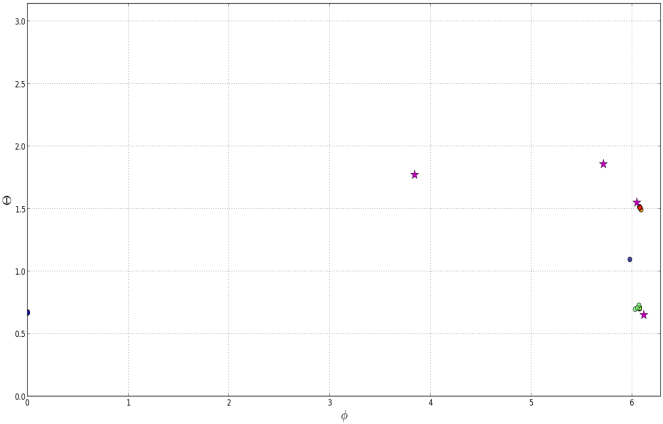
Dataset 2

60 nHz high pass filter



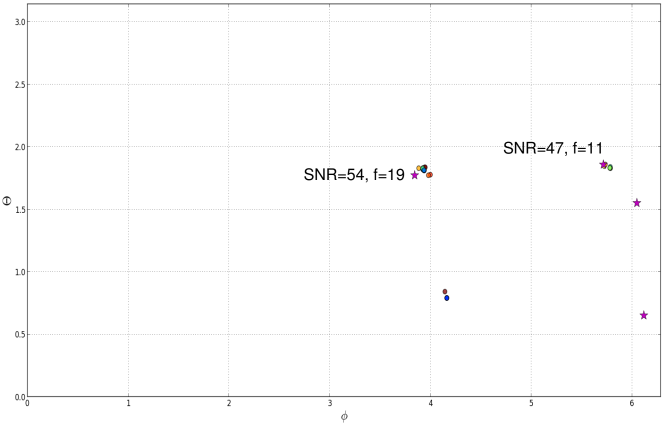
Dataset 2

40 nHz high pass filter



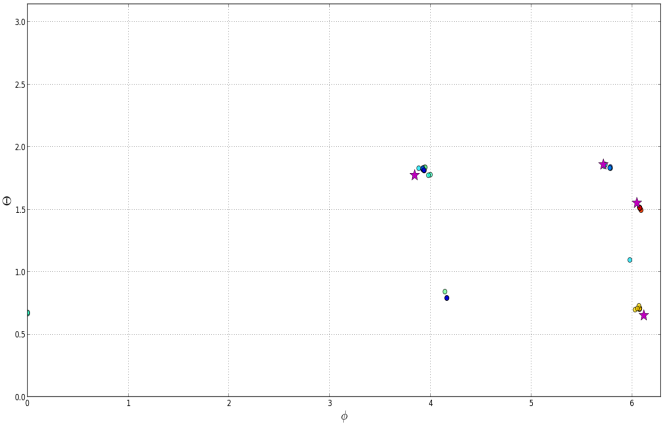
Dataset 2

20 nHz high pass filter



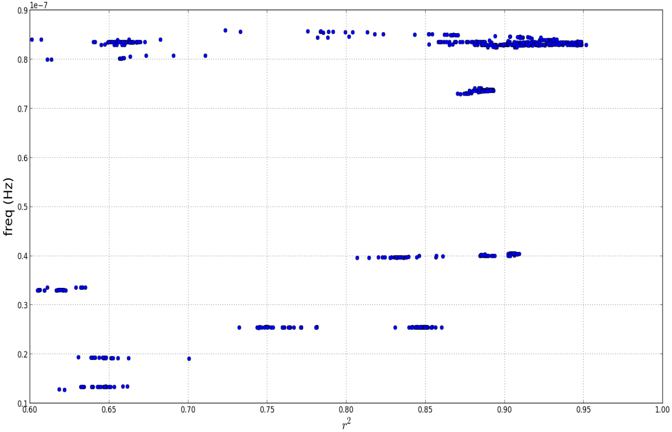
Dataset 2

Combined plot



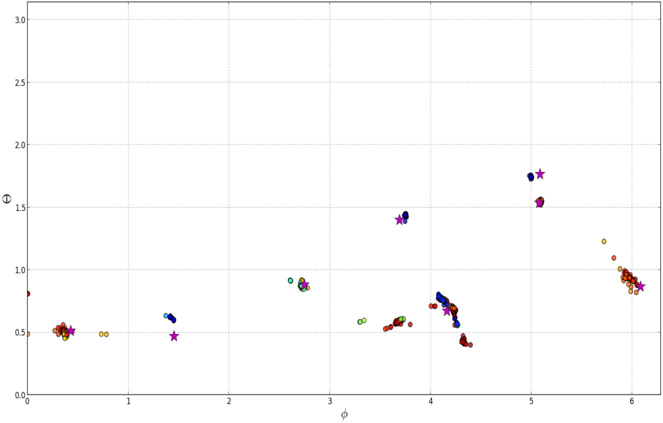
Dataset 3

The sources were weak: we did use filters. Many source (8).



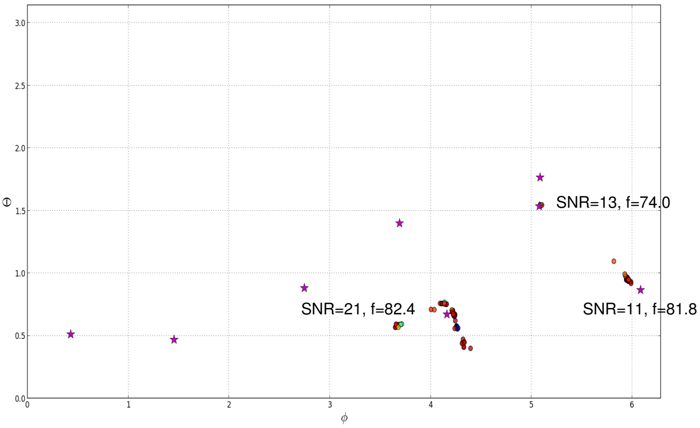
Dataset 3

The sources were weak: we did use filters.



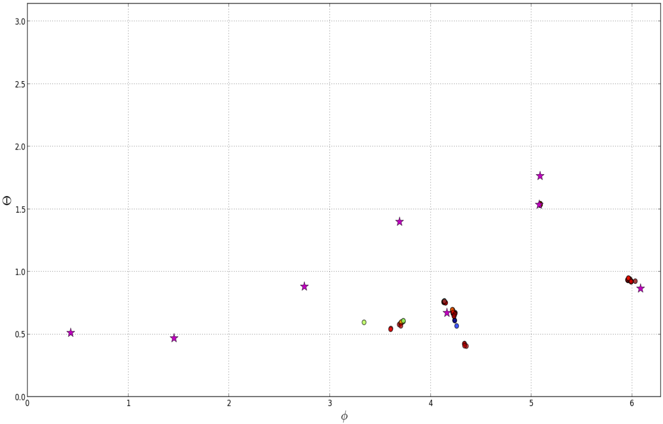
Dataset 3

140 nHz high pass filter



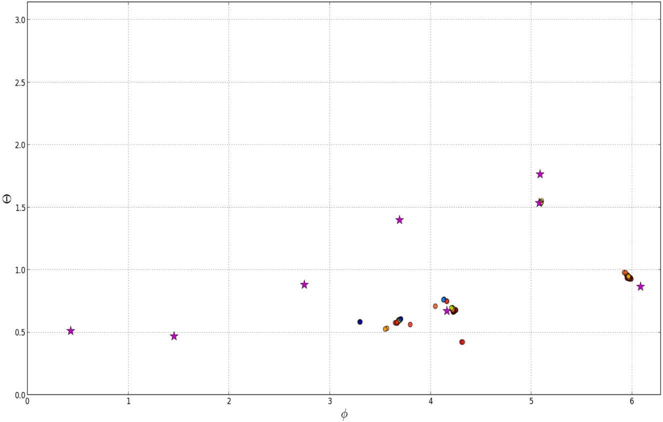
Dataset 3

120 nHz high pass filter



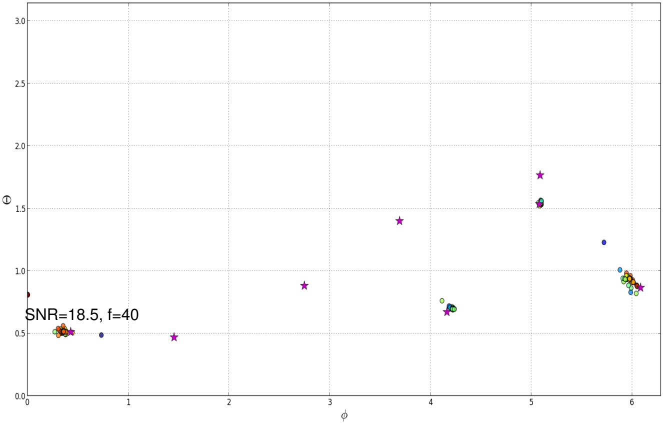
Dataset 3

100 nHz high pass filter



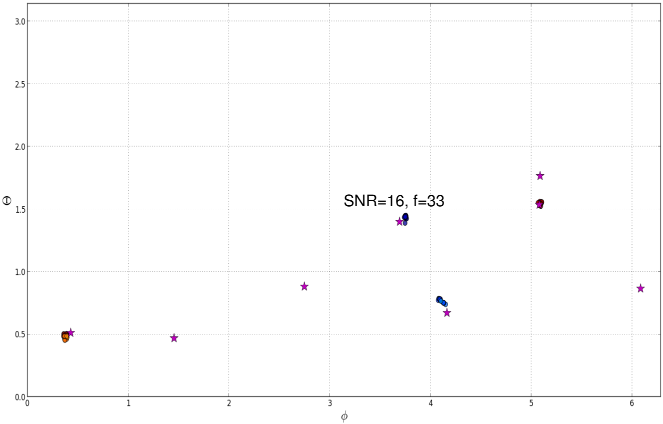
Dataset 3

80 nHz high pass filter



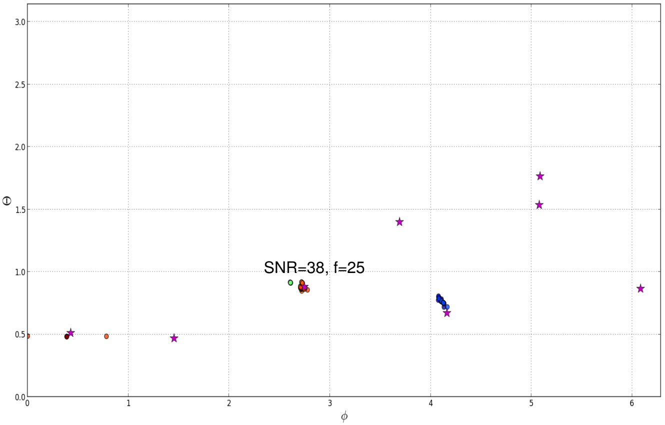
Dataset 3

60 nHz high pass filter



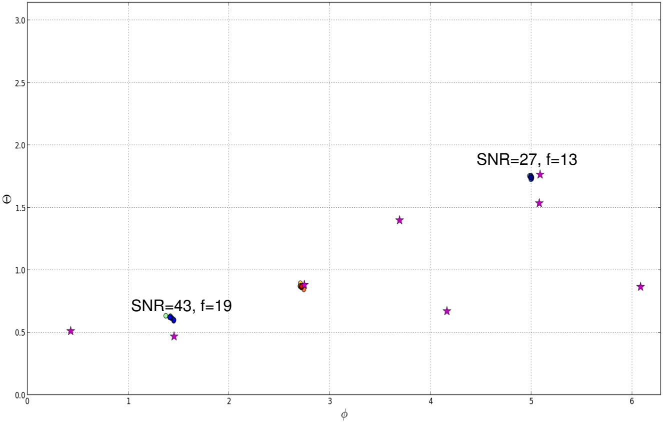
Dataset 3

40 nHz high pass filter



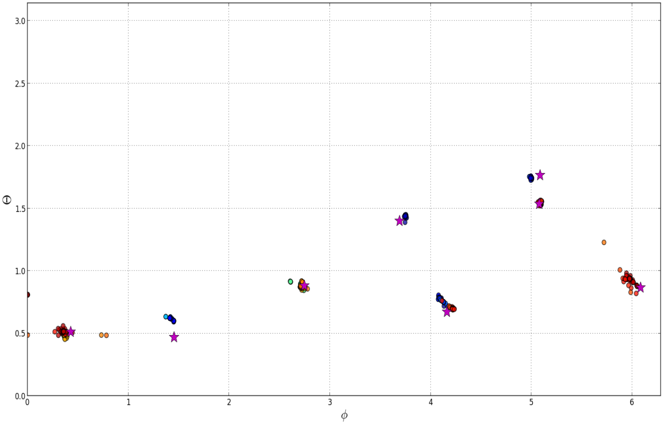
Dataset 3

20 nHz high pass filter



Dataset 3

Combined plot



Summary

- We have presented method for searching for multiple SMBH binaries with PTA
- We have used simplified data and high SNR sources (will make it more realistic in the next steps)
- We use detection statistic based on the “earth term” in the response only and treat “pulsar term” as source of non-Gaussian non-stationary noise.
- We have used multimodal Genetic Algorithm for search and combined the results of the search with consistency check to eliminate the “pulsar-term-generated” candidates.
- We have done three (semi-)blind analysis with a good recovery of the source parameters.

Summary

- We have presented method for searching for multiple SMBH binaries with PTA
- We have used simplified data and high SNR sources (will make it more realistic in the next steps)
- We use detection statistic based on the "earth term" in the response only and treat "pulsar term" as source of non-Gaussian non-stationary noise.
- We have used multimodal Genetic Algorithm for search and combined the results of the search with consistency check to eliminate the "pulsar-term-generated" candidates.
- We have done three (semi-)blind analysis with a good recovery of the source parameters.

Summary

- We have presented method for searching for multiple SMBH binaries with PTA
- We have used simplified data and high SNR sources (will make it more realistic in the next steps)
- We use detection statistic based on the “earth term” in the response only and treat “pulsar term” as source of non-Gaussian non-stationary noise.
- We have used multimodal Genetic Algorithm for search and combined the results of the search with consistency check to eliminate the “pulsar-term-generated” candidates.
- We have done three (semi-)blind analysis with a good recovery of the source parameters.

Summary

- We have presented method for searching for multiple SMBH binaries with PTA
- We have used simplified data and high SNR sources (will make it more realistic in the next steps)
- We use detection statistic based on the “earth term” in the response only and treat “pulsar term” as source of non-Gaussian non-stationary noise.
- We have used multimodal Genetic Algorithm for search and combined the results of the search with consistency check to eliminate the “pulsar-term-generated” candidates.
- We have done three (semi-)blind analysis with a good recovery of the source parameters.

Summary

- We have presented method for searching for multiple SMBH binaries with PTA
- We have used simplified data and high SNR sources (will make it more realistic in the next steps)
- We use detection statistic based on the "earth term" in the response only and treat "pulsar term" as source of non-Gaussian non-stationary noise.
- We have used multimodal Genetic Algorithm for search and combined the results of the search with consistency check to eliminate the "pulsar-term-generated" candidates.
- We have done three (semi-)blind analysis with a good recovery of the source parameters.