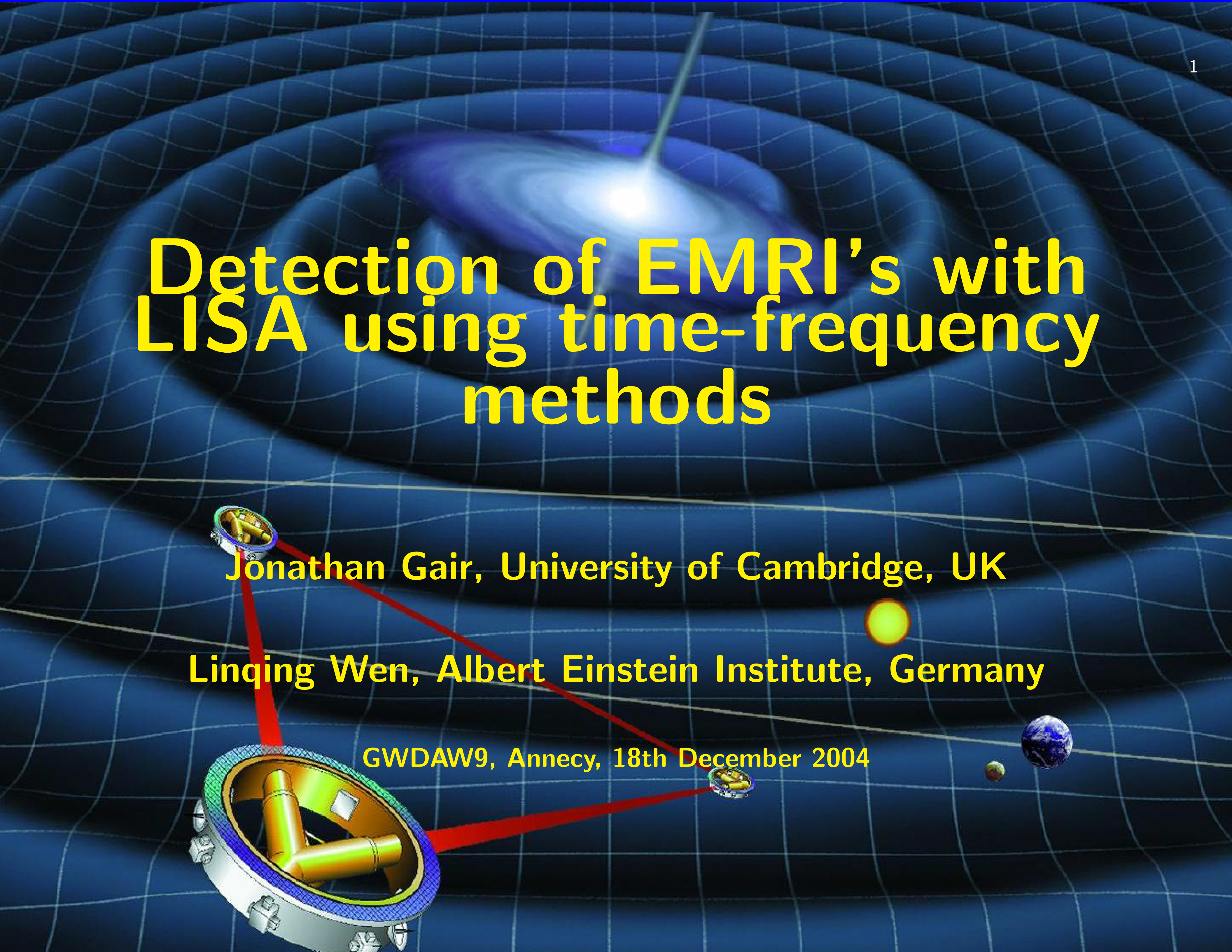


Detection of EMRI's with LISA using time-frequency methods

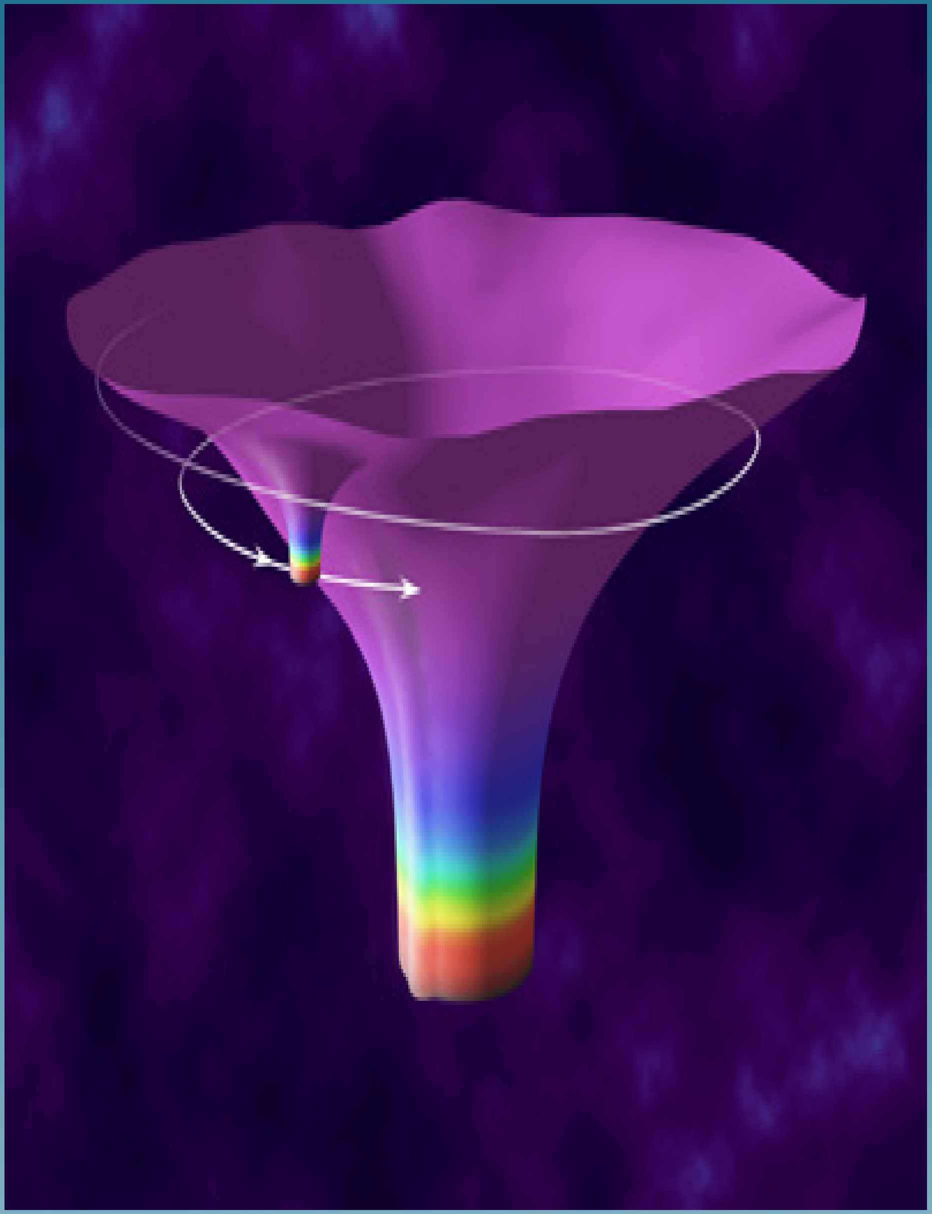
Jonathan Gair, University of Cambridge, UK

Linqing Wen, Albert Einstein Institute, Germany

GWDAW9, Annecy, 18th December 2004

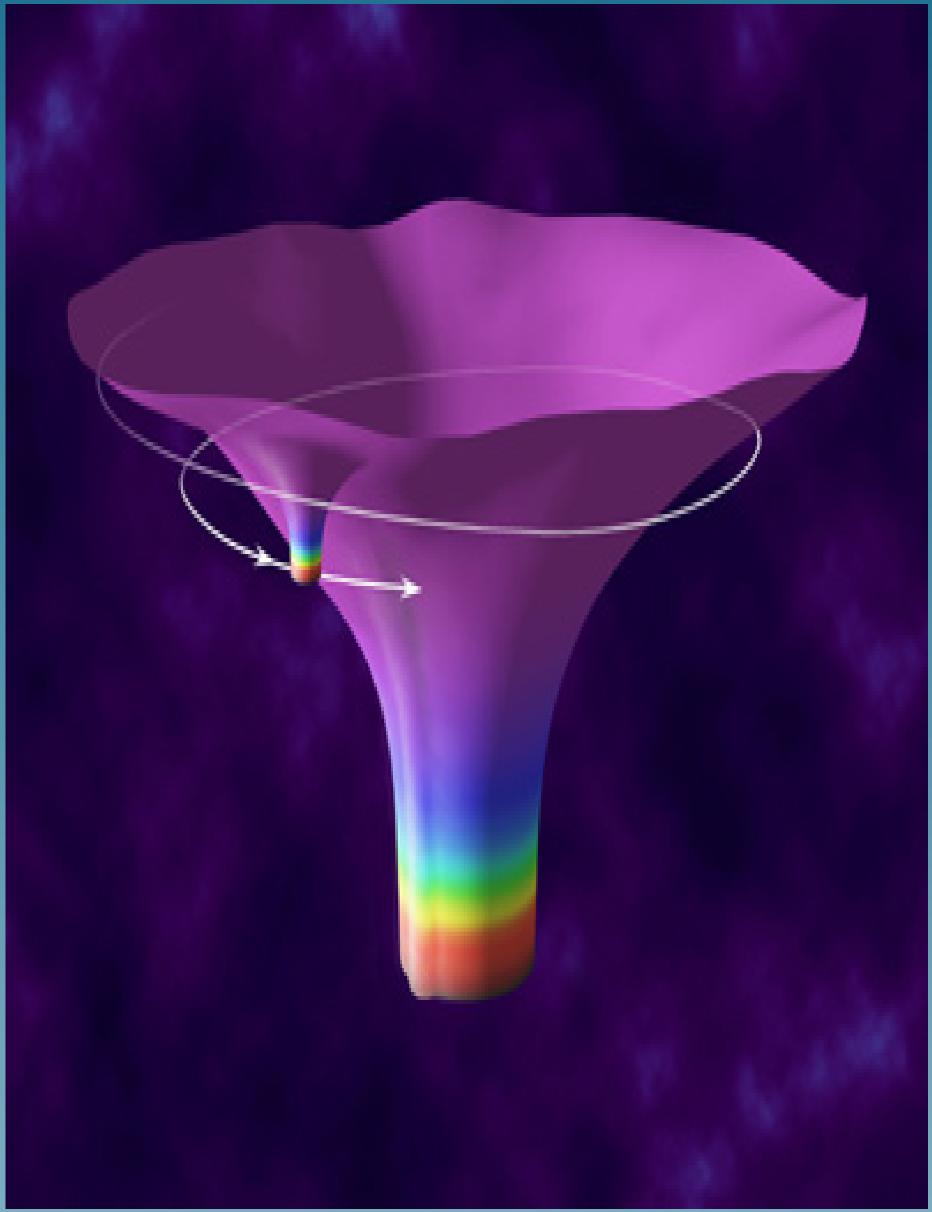
Extreme mass ratio inspirals

- Inspiral of a compact body into a supermassive black hole.



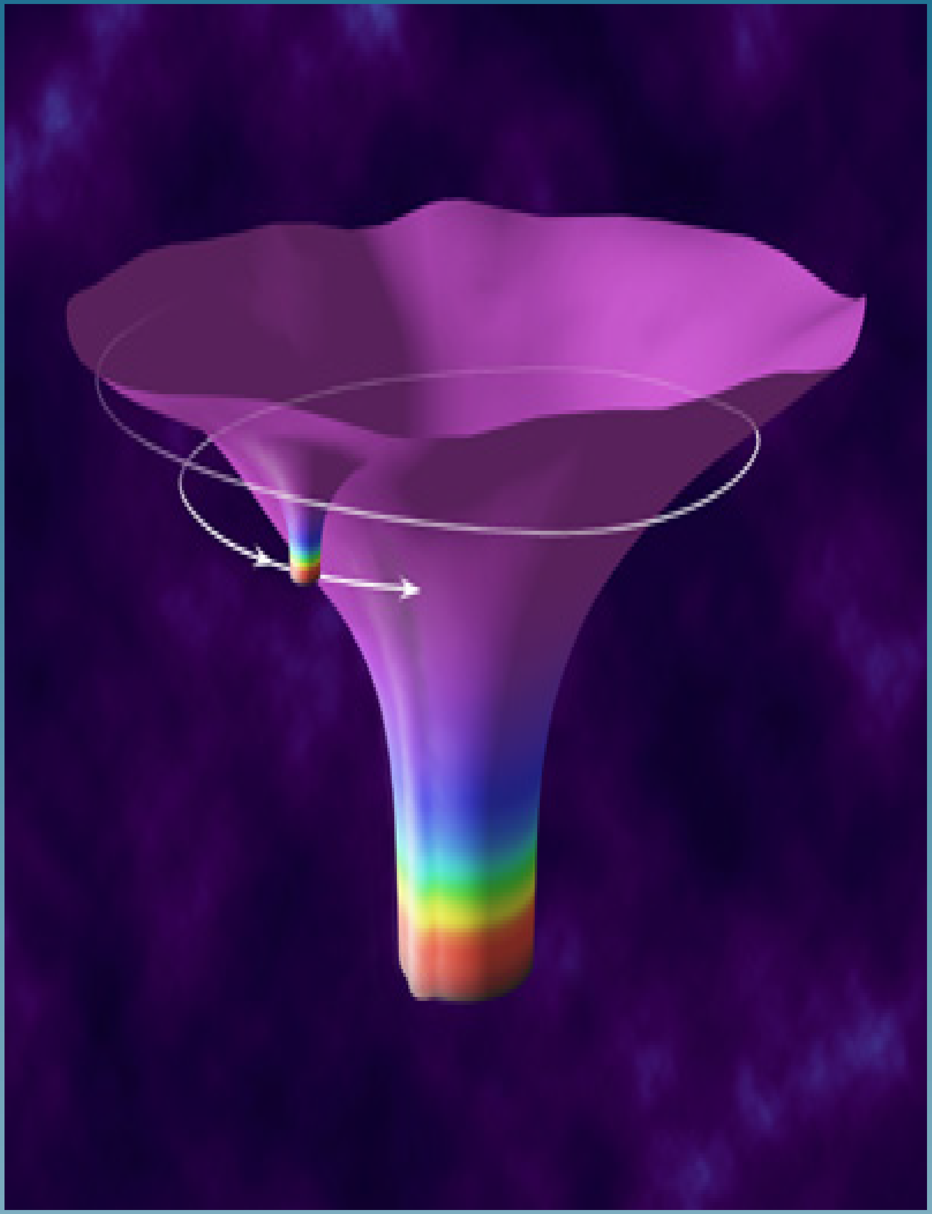
Extreme mass ratio inspirals

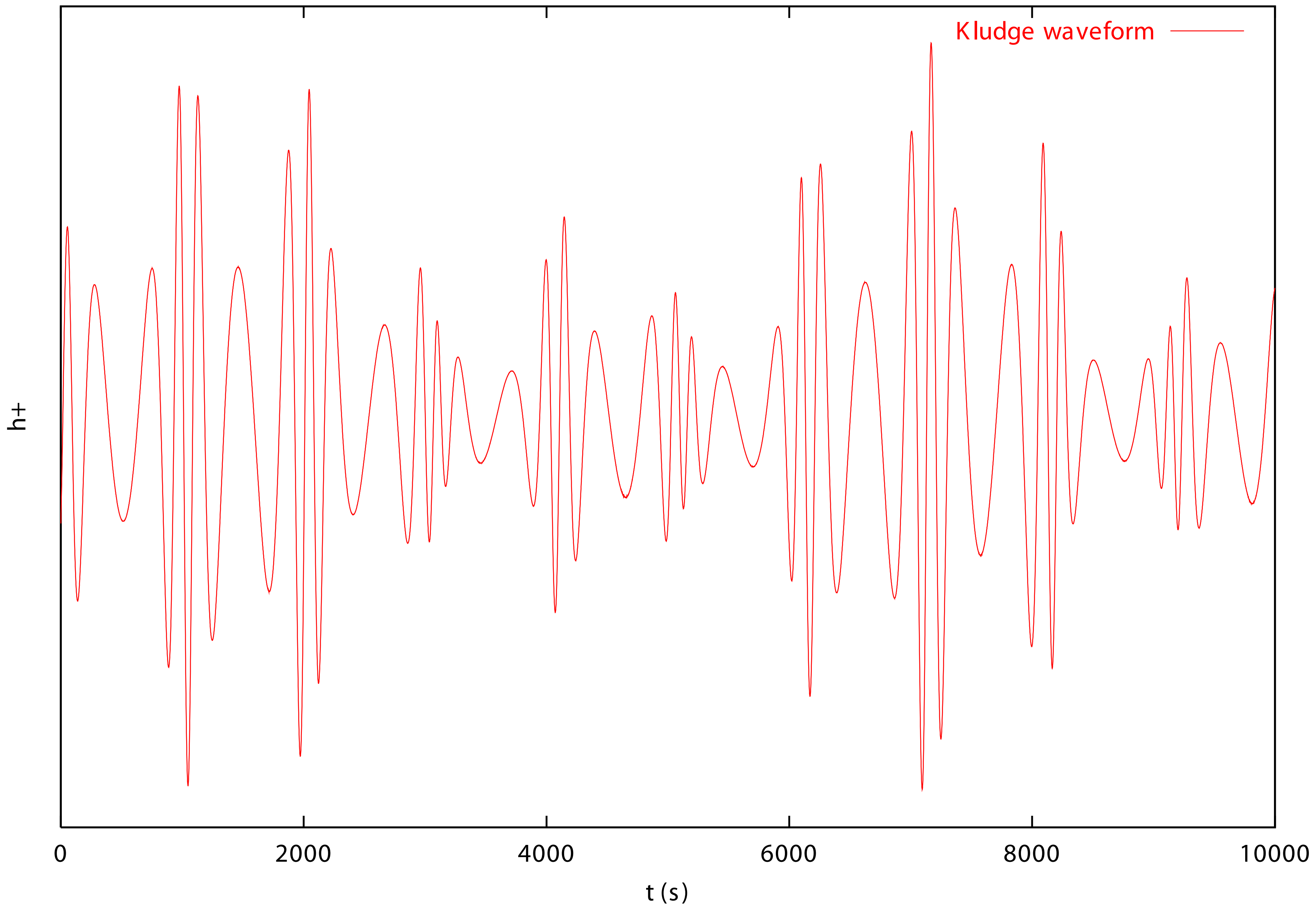
- Inspiral of a compact body into a supermassive black hole.
- Inspirals radiate in the LISA band for $M \sim 10^5 - 10^7 M_\odot$.



Extreme mass ratio inspirals

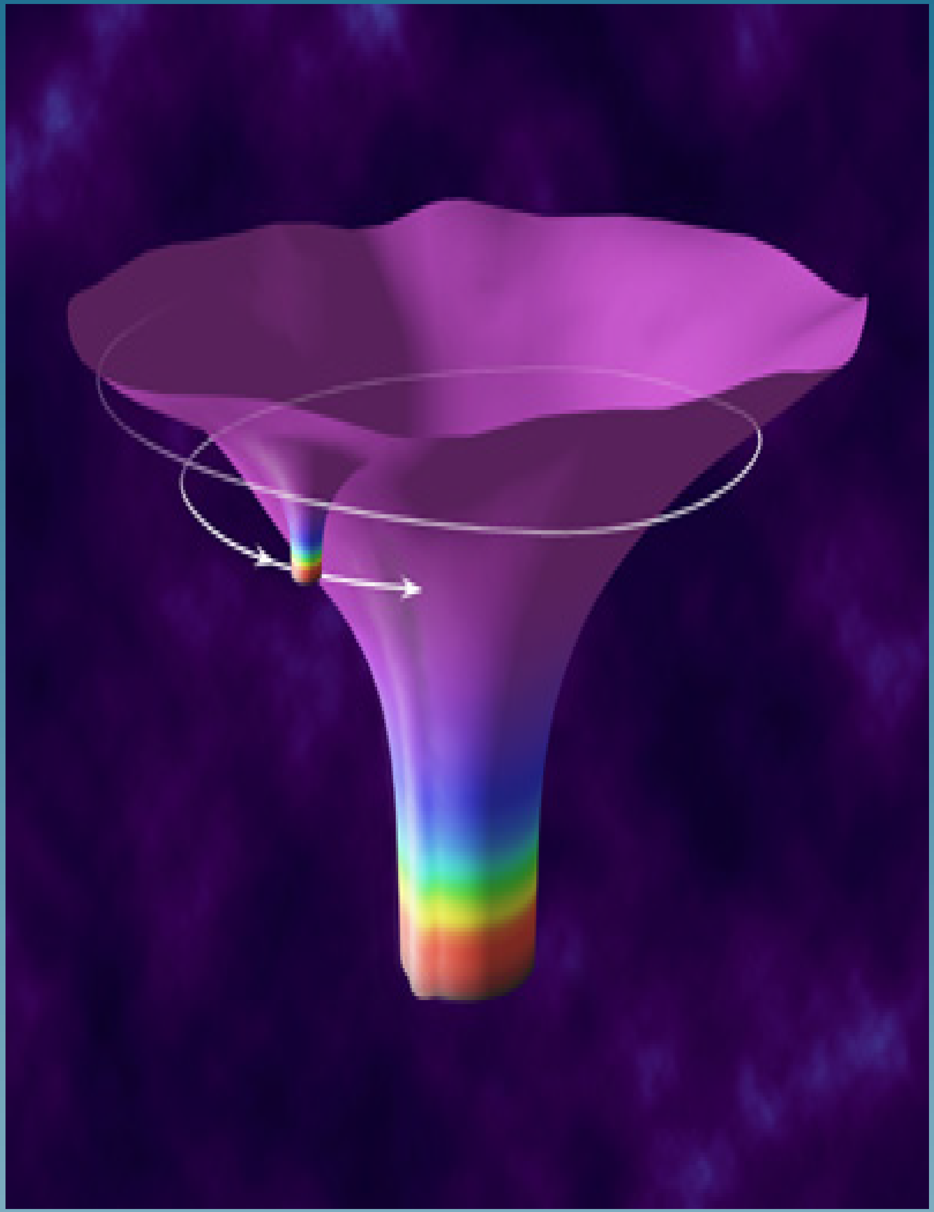
- Inspiral of a compact body into a supermassive black hole.
- Inspirals radiate in the LISA band for $M \sim 10^5 - 10^7 M_\odot$.
- Complicated gravitational waveforms encode a map of the spacetime geometry around spinning black holes. Uncode this map to probe spacetime structure - "holiodesy".




Kludge waveform
h+
0
2000
4000
6000
8000
10000
t (s)

Extreme mass ratio inspirals

- Inspiral of a compact body into a supermassive black hole.
- Inspirals radiate in the LISA band for $M \sim 10^5 - 10^7 M_\odot$.
- Complicated gravitational waveforms encode a map of the spacetime geometry around spinning black holes. Uncode this map to probe spacetime structure - "holiodesy".
- The potential scientific impact has made detection of a significant number of EMRI events a key LISA science requirement.



Detection of EMRI's

- The EMRI parameter space is very large - waveforms depend on 17 different parameters - and expect to observe large numbers of cycles ($\sim 10^5$). A search based on fully coherent matched filtering is thus computationally impossible.

Detection of EMRI's

- The EMRI parameter space is very large - waveforms depend on 17 different parameters - and expect to observe large numbers of cycles ($\sim 10^5$). A search based on fully coherent matched filtering is thus computationally impossible.
- Detection could be achieved using a mixed coherent/incoherent search. A first cut analysis (Gair *et al.* 2004) indicated
  - ⋆ Threshold SNR is increased from ~ 15 (fully coherent) to ~ 35, but events can still be detected to $z \sim 1$.

Detection of EMRI's

- The EMRI parameter space is very large - waveforms depend on 17 different parameters - and expect to observe large numbers of cycles ($\sim 10^5$). A search based on fully coherent matched filtering is thus computationally impossible.

- Detection could be achieved using a mixed coherent/incoherent search. A first cut analysis (Gair *et al.* 2004) indicated
  - ⋆ Threshold SNR is increased from ~ 15 (fully coherent) to ~ 35, but events can still be detected to $z \sim 1$.
  - ⋆ Between ~ 100 and ~ 1000s of events could be observed, depending on the intrinsic astrophysical rate.

Detection of EMRI's

- The EMRI parameter space is very large - waveforms depend on 17 different parameters - and expect to observe large numbers of cycles ($\sim 10^5$). A search based on fully coherent matched filtering is thus computationally impossible.
- Detection could be achieved using a mixed coherent/incoherent search. A first cut analysis (Gair *et al.* 2004) indicated
  - ⋆ Threshold SNR is increased from ~ 15 (fully coherent) to ~ 35, but events can still be detected to $z \sim 1$.
  - ⋆ Between ~ 100 and ~ 1000s of events could be observed, depending on the intrinsic astrophysical rate.
- The semi-coherent search is still computationally expensive and makes maximum use of available resources (analysis on a ~ 50 Teraflop computer in real time).

Detection of EMRI's

- The EMRI parameter space is very large - waveforms depend on 17 different parameters - and expect to observe large numbers of cycles ($\sim 10^5$). A search based on fully coherent matched filtering is thus computationally impossible.

- Detection could be achieved using a mixed coherent/incoherent search. A first cut analysis (Gair *et al.* 2004) indicated
  - ⋆ Threshold SNR is increased from ~ 15 (fully coherent) to ~ 35, but events can still be detected to $z \sim 1$.
  - ⋆ Between ~ 100 and ~ 1000s of events could be observed, depending on the intrinsic astrophysical rate.

- The semi-coherent search is still computationally expensive and makes maximum use of available resources (analysis on a ~ 50 Teraflop computer in real time).

- Valuable to explore other search techniques to use in conjunction with matched filtering algorithms.

Time-frequency analysis I

- A time-frequency analysis is one possible alternative approach to EMRI detection. To investigate the feasibility of such methods, we have scoped out a very simple time-frequency algorithm. It is similar to the excess power statistic used in LIGO data analysis.

Time-frequency analysis I

- A time-frequency analysis is one possible alternative approach to EMRI detection. To investigate the feasibility of such methods, we have scoped out a very simple time-frequency algorithm. It is similar to the excess power statistic used in LIGO data analysis.
- Make simplifying assumptions in first cut - assume a clean data stream (i.e., no white dwarf binaries) and that we are searching for a single EMRI event.

Time-frequency analysis I

- A time-frequency analysis is one possible alternative approach to EMRI detection. To investigate the feasibility of such methods, we have scoped out a very simple time-frequency algorithm. It is similar to the excess power statistic used in LIGO data analysis.
- Make simplifying assumptions in first cut - assume a clean data stream (i.e., no white dwarf binaries) and that we are searching for a single EMRI event.
- Analyse LISA data as follows
  - ⋆ Divide data in time into sections of length T (~ 2 weeks). Compute SFT's and construct normalised power at each time i and frequency j

$$P(i,j) = 4\, dt^2\, df\, \frac{|s^i_j|^2}{S_h(f_j)}$$

Time-frequency analysis I

- A time-frequency analysis is one possible alternative approach to EMRI detection. To investigate the feasibility of such methods, we have scoped out a very simple time-frequency algorithm. It is similar to the excess power statistic used in LIGO data analysis.

- Make simplifying assumptions in first cut - assume a clean data stream (i.e., no white dwarf binaries) and that we are searching for a single EMRI event.

- Analyse LISA data as follows
  - ⋆ Divide data in time into sections of length T (~ 2 weeks). Compute SFT's and construct normalised power at each time i and frequency j

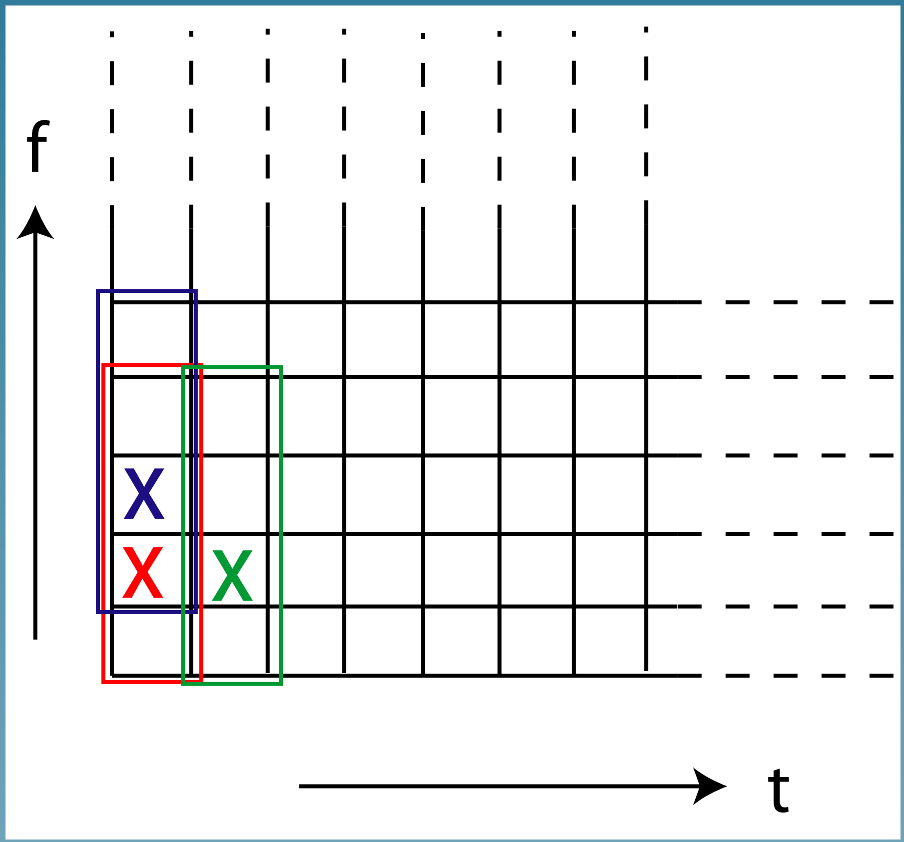
$$P(i,j) = 4\, dt^2\, df\, \frac{|s^i_j|^2}{S_h(f_j)}$$

  - ⋆ Combine power from two independent data streams.

Time-frequency analysis II

- Search for areas of high power density by binning using rectangular grids in the t-f plane.

$$\rho(i,j) = \sum_{k=-\frac{n_t}{2}}^{\frac{n_t}{2}} \sum_{l=-\frac{n_f}{2}}^{\frac{n_f}{2}} \frac{P(i+k, j+l)}{n_t\, n_f}$$

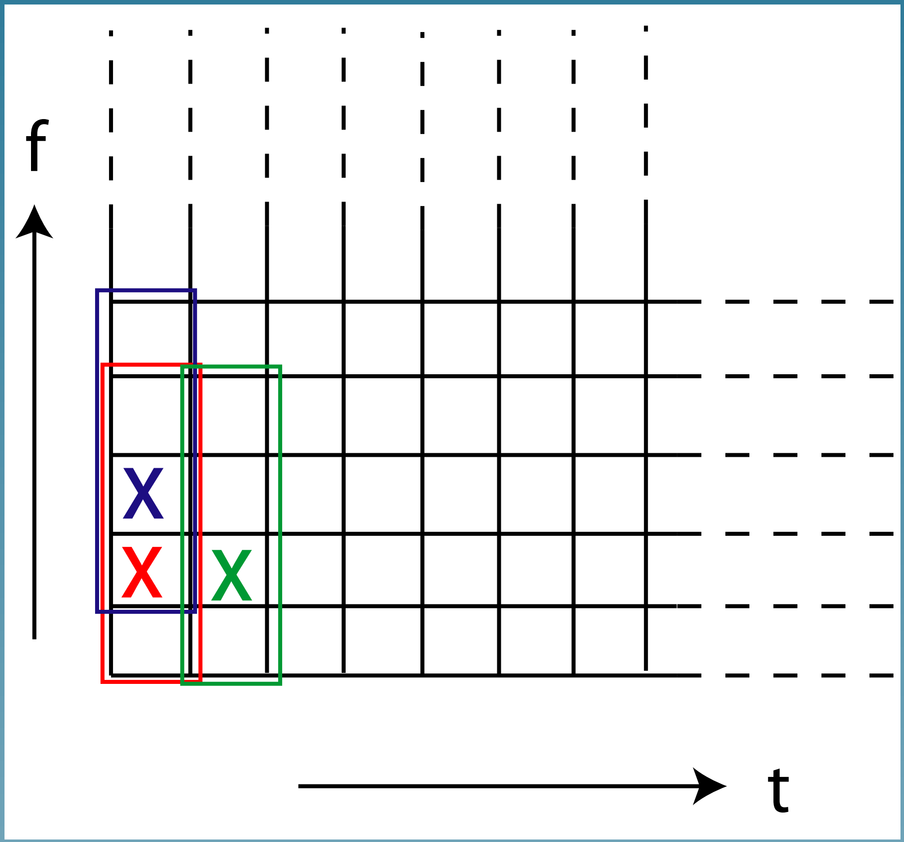


Time-frequency analysis II

- Search for areas of high power density by binning using rectangular grids in the t-f plane.

$$\rho(i,j) = \sum_{k=-\frac{n_t}{2}}^{\frac{n_t}{2}} \sum_{l=-\frac{n_f}{2}}^{\frac{n_f}{2}} \frac{P(i+k,j+l)}{n_t\, n_f}$$

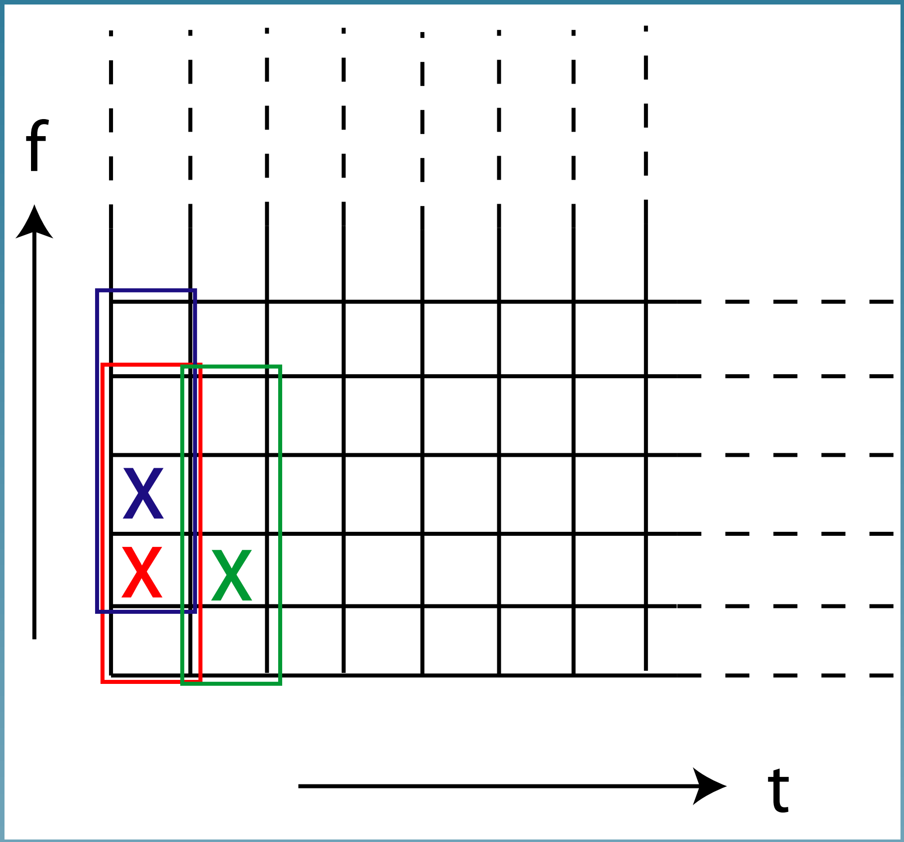
- Could use knowledge of source waveforms to restrict choice of bin size, but here we consider a blind search over a range of bin sizes.



Time-frequency analysis II

- Search for areas of high power density by binning using rectangular grids in the t-f plane.

$$\rho(i,j) = \sum_{k=-\frac{n_t}{2}}^{\frac{n_t}{2}} \sum_{l=-\frac{n_f}{2}}^{\frac{n_f}{2}} \frac{P(i+k, j+l)}{n_t\, n_f}$$



- Could use knowledge of source waveforms to restrict choice of bin size, but here we consider a blind search over a range of bin sizes.
- For each bin size, find false alarm probability of loudest excess in the t-f plane. Minimize over all bin sizes.

Search statistic

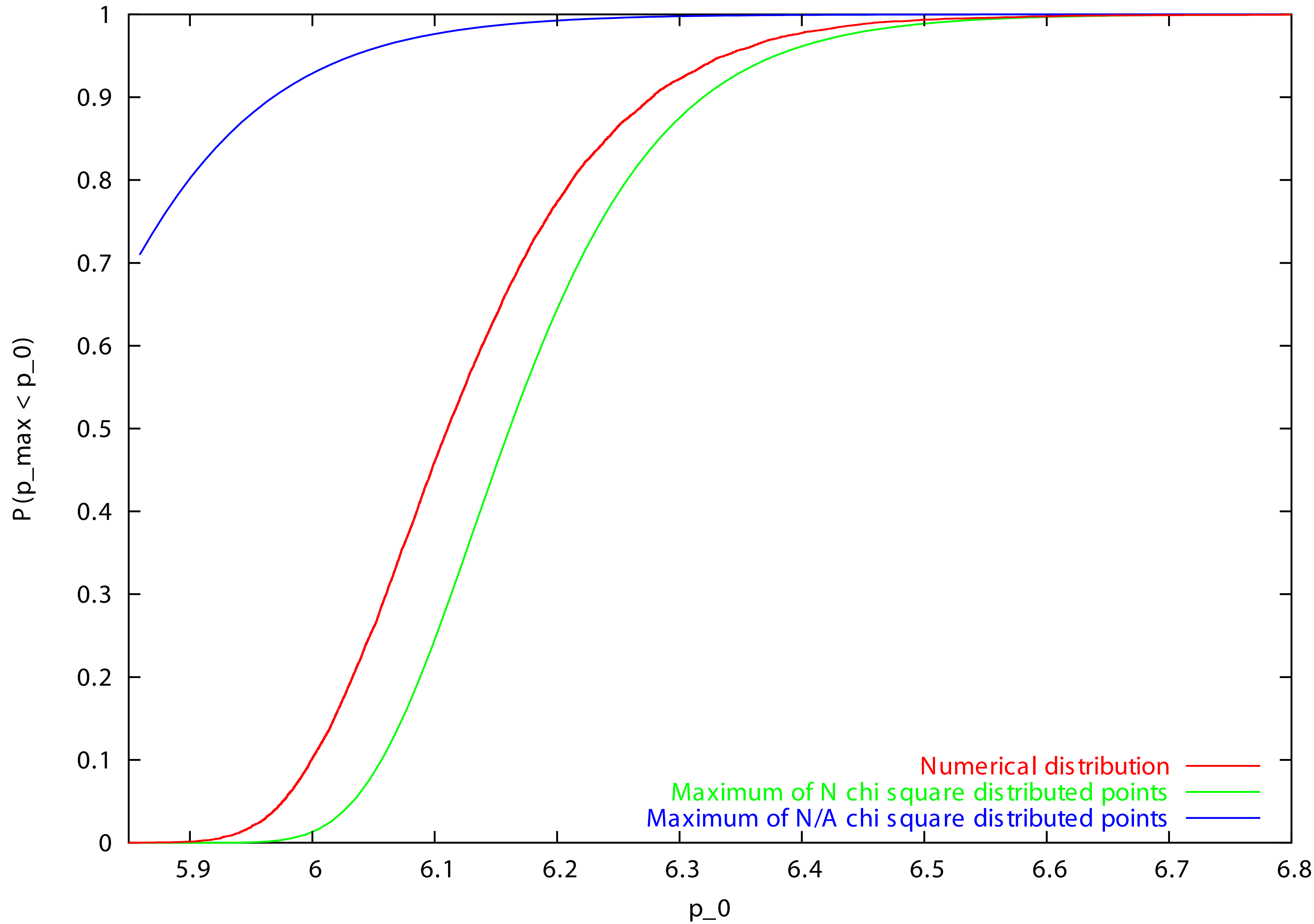
- For a specified bin size and in the absence of a signal, the power in a given pixel, $p = A\,\rho$, follows a chi-square distribution with $4\,A$ degrees of freedom, $P_{\chi^2_{4A}}(p < p_0)$.

Search statistic

- For a specified bin size and in the absence of a signal, the power in a given pixel, $p = A\,\rho$, follows a chi-square distribution with $4\,A$ degrees of freedom, $P_{\chi^2_{4A}}(p < p_0)$.
- If we drew l independent samples from this distribution, the maximum would be distributed as $P(p_{max} < p_0) = \left[P_{\chi^2_{4A}}(p < p_0) \right]^l$.

Search statistic

- For a specified bin size and in the absence of a signal, the power in a given pixel, $p = A\,\rho$, follows a chi-square distribution with $4\,A$ degrees of freedom, $P_{\chi^2_{4A}}(p < p_0)$.
- If we drew l independent samples from this distribution, the maximum would be distributed as $P(p_{max} < p_0) = \left[P_{\chi^2_{4A}}(p < p_0)\right]^l$.
- In the binned t-f plane, overlap between bins makes statistics more complicated, but expect to have effectively $N/A < l < N$. Use Monte Carlo simulations to compute $P_i(p_0) = P((p_{max})_i < p_0)$ numerically for each bin type i.



P(p_max < p_0)
p_0
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
5.9
6
6.1
6.2
6.3
6.4
6.5
6.6
6.7
6.8
Numerical distribution
Maximum of N chi square distributed points
Maximum of N/A chi square distributed points

Search statistic

- For a specified bin size and in the absence of a signal, the power in a given pixel, $p = A\,\rho$, follows a chi-square distribution with $4\,A$ degrees of freedom, $P_{\chi^2_{4A}}(p < p_0)$.
- If we drew l independent samples from this distribution, the maximum would be distributed as $P(p_{max} < p_0) = \left[P_{\chi^2_{4A}}(p < p_0)\right]^l$.
- In the binned t-f plane, overlap between bins makes statistics more complicated, but expect to have effectively $N/A < l < N$. Use Monte Carlo simulations to compute $P_i(p_0) = P((p_{max})_i < p_0)$ numerically for each bin type i.
- Also use Monte Carlo simulation to compute distribution of final search statistic, $\mathrm{FAP}_{min} = \min\left\{1 - P_i((p_{max})_i^{obs})\right\}$. If all M bin sizes were independent, then $P(\mathrm{FAP}_{min} < Y) = 1-(1-Y)^M$. Simulations indicate we should set a threshold $\mathrm{FAP}_{min} \approx 10^{-4}$ to give the search an overall false alarm probability of 1%.

Test case

- To estimate the effectiveness of this method, we used a mock LISA data stream containing a "typical" EMRI event with parameters
  - ⋆ Inspiral of $m = 10M_{\odot}$ BH into a $M = 10^6 M_{\odot}$ SMBH with spin $a = 0.8M$.

Test case

- To estimate the effectiveness of this method, we used a mock LISA data stream containing a "typical" EMRI event with parameters
  - ⋆ Inspiral of $m = 10 M_\odot$ BH into a $M = 10^6 M_\odot$ SMBH with spin $a = 0.8M$.
  - ⋆ Used waveform over the last three years of inspiral, with eccentricity $e = 0.4$, periapse $r_p \approx 11M$ and inclination, $\iota = 45^{\mathrm{o}}$ at the start of observation.

Test case

- To estimate the effectiveness of this method, we used a mock LISA data stream containing a "typical" EMRI event with parameters
  - ⋆ Inspiral of $m = 10M_{\odot}$ BH into a $M = 10^6 M_{\odot}$ SMBH with spin $a = 0.8M$.
  - ⋆ Used waveform over the last three years of inspiral, with eccentricity $e = 0.4$, periapse $r_p \approx 11M$ and inclination, $\iota = 45^{\text{o}}$ at the start of observation.
  - ⋆ Source was placed at a fixed, non-special, point on the sky at distances of $d = 0.5$, 1, 1.4 and 2 Gpc.

Test case

- To estimate the effectiveness of this method, we used a mock LISA data stream containing a "typical" EMRI event with parameters
  - ⋆ Inspiral of $m = 10M_\odot$ BH into a $M = 10^6 M_\odot$ SMBH with spin $a = 0.8M$.
  - ⋆ Used waveform over the last three years of inspiral, with eccentricity $e = 0.4$, periapse $r_p \approx 11M$ and inclination, $\iota = 45^o$ at the start of observation.
  - ⋆ Source was placed at a fixed, non-special, point on the sky at distances of $d = 0.5$, 1, 1.4 and 2 Gpc.
- The inspiral waveform was generated using a numerical kludge technique, including modulations induced by the detector motion.

Test case

- To estimate the effectiveness of this method, we used a mock LISA data stream containing a "typical" EMRI event with parameters
  - ⋆ Inspiral of $m = 10M_\odot$ BH into a $M = 10^6 M_\odot$ SMBH with spin $a = 0.8M$.
  - ⋆ Used waveform over the last three years of inspiral, with eccentricity $e = 0.4$, periapse $r_p \approx 11M$ and inclination, $\iota = 45^o$ at the start of observation.
  - ⋆ Source was placed at a fixed, non-special, point on the sky at distances of $d = 0.5$, 1, 1.4 and 2 Gpc.

- The inspiral waveform was generated using a numerical kludge technique, including modulations induced by the detector motion.

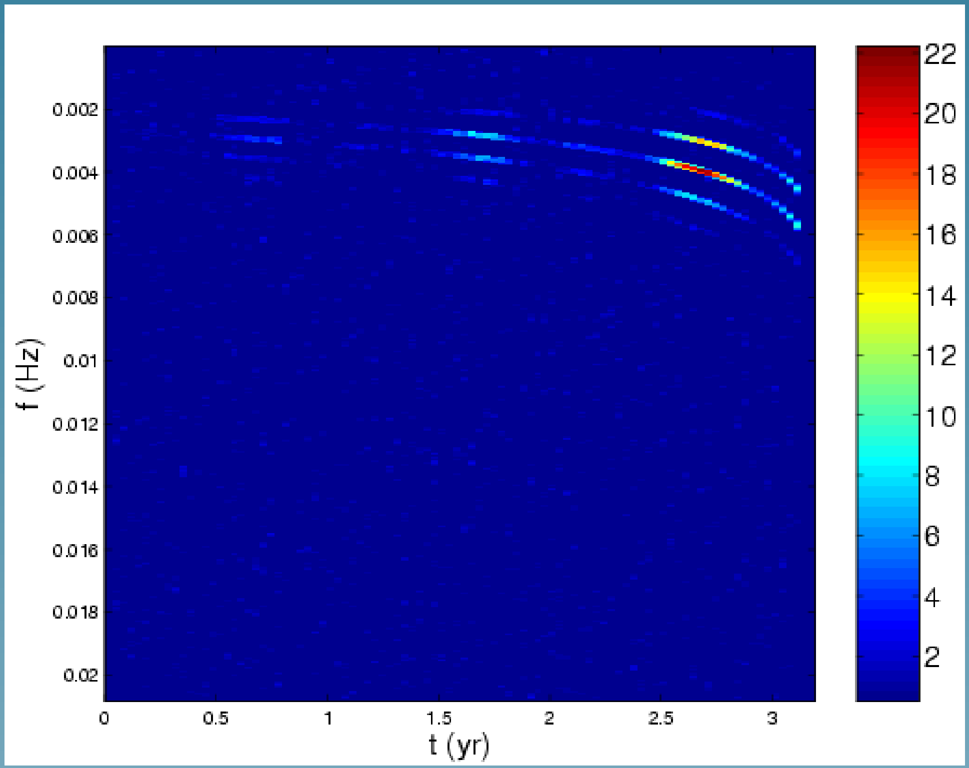
- This source would have a matched filtering SNR of ~ 120 at 1 Gpc and could be detected out to $z > 1$ using the semi-coherent technique outlined previously.

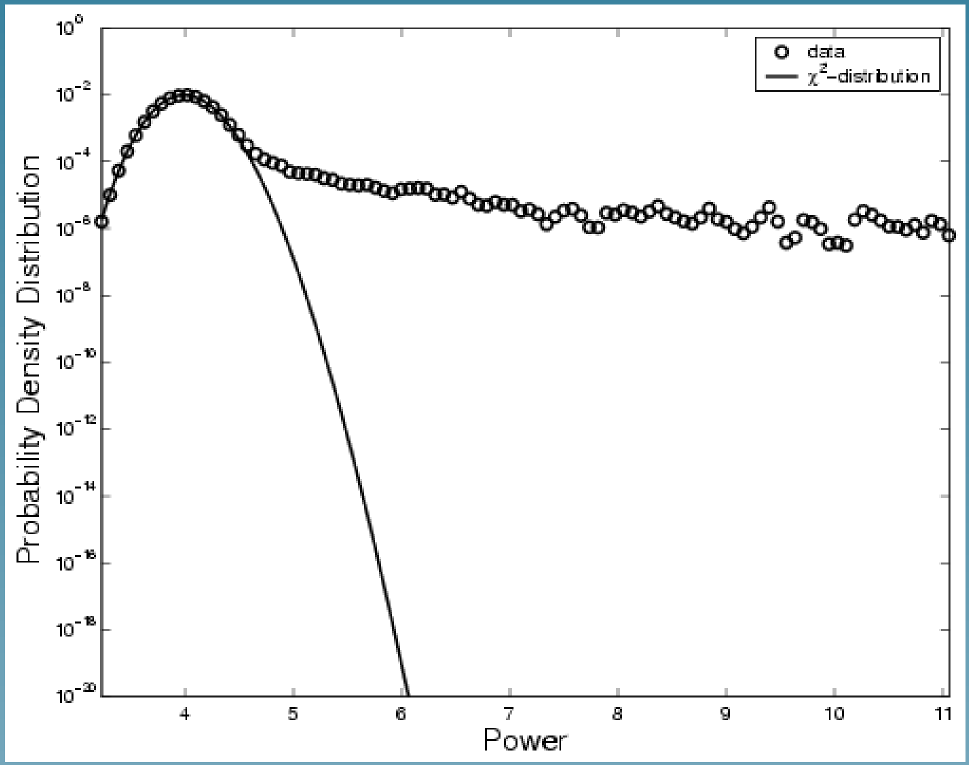
Test case

- To estimate the effectiveness of this method, we used a mock LISA data stream containing a "typical" EMRI event with parameters
  - ⋆ Inspiral of $m = 10M_\odot$ BH into a $M = 10^6 M_\odot$ SMBH with spin $a = 0.8M$.
  - ⋆ Used waveform over the last three years of inspiral, with eccentricity $e = 0.4$, periapse $r_p \approx 11M$ and inclination, $\iota = 45^o$ at the start of observation.
  - ⋆ Source was placed at a fixed, non-special, point on the sky at distances of $d = 0.5$, 1, 1.4 and 2 Gpc.
- The inspiral waveform was generated using a numerical kludge technique, including modulations induced by the detector motion.
- This source would have a matched filtering SNR of ~ 120 at 1 Gpc and could be detected out to $z > 1$ using the semi-coherent technique outlined previously.
- Using the time-frequency analysis, the source could be detected out to $d \sim 2$ Gpc, giving a maximum event rate ~ 100 over three years.

Results - source at d = 0.5 Gpc

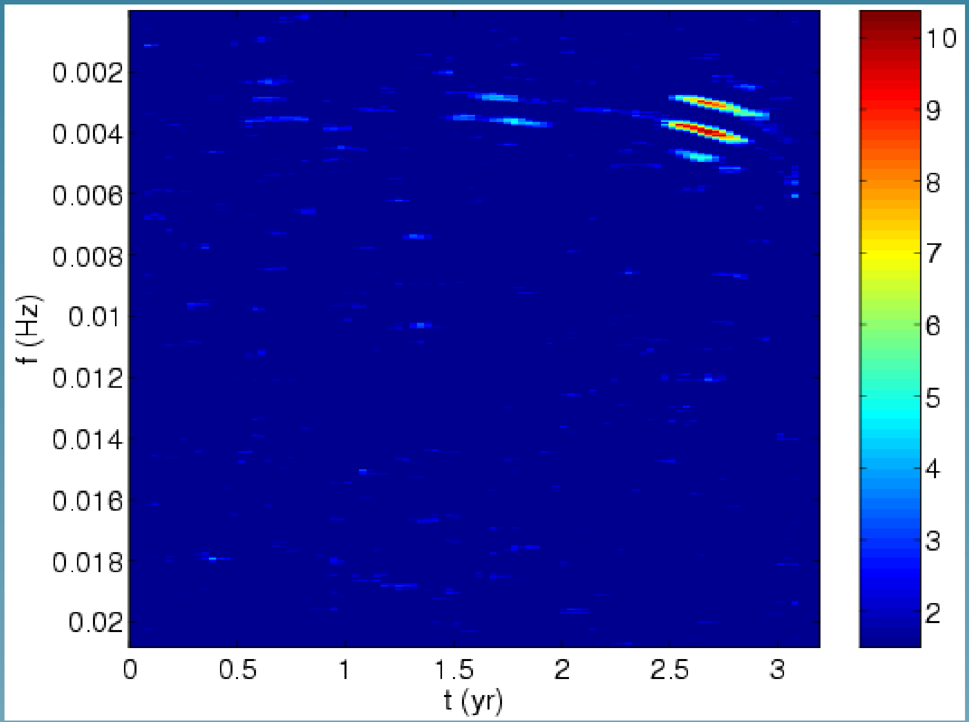
- Nearest likely event is detected with high confidence. Expect 0-2 events with $d < 0.5$ Gpc over three years.

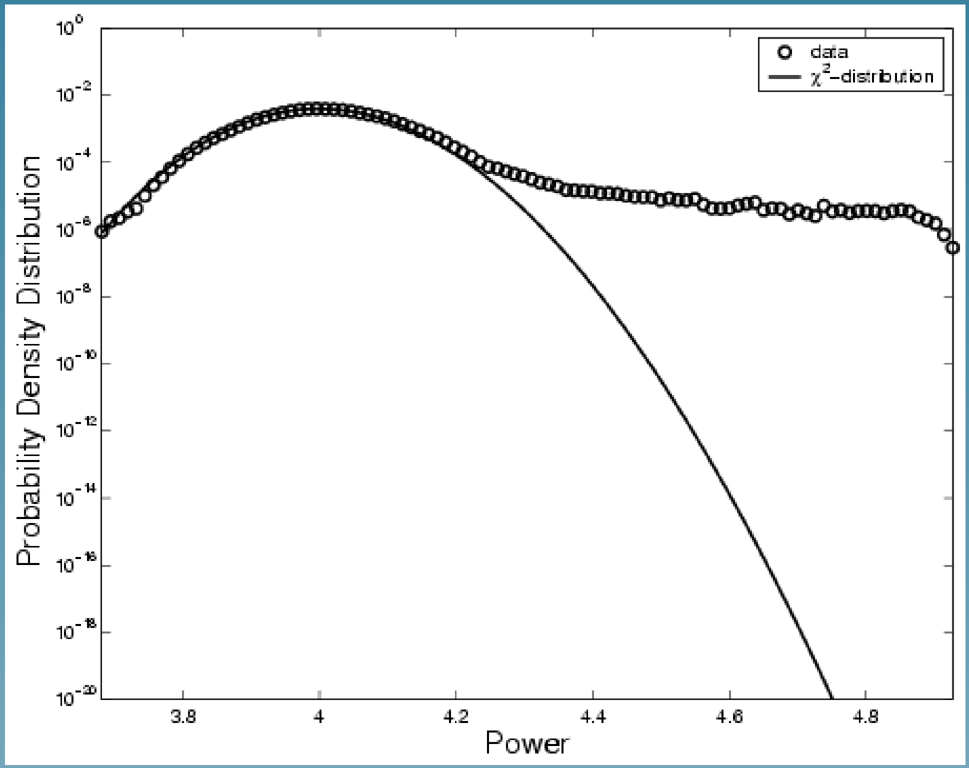




Results - source at d = 1 Gpc

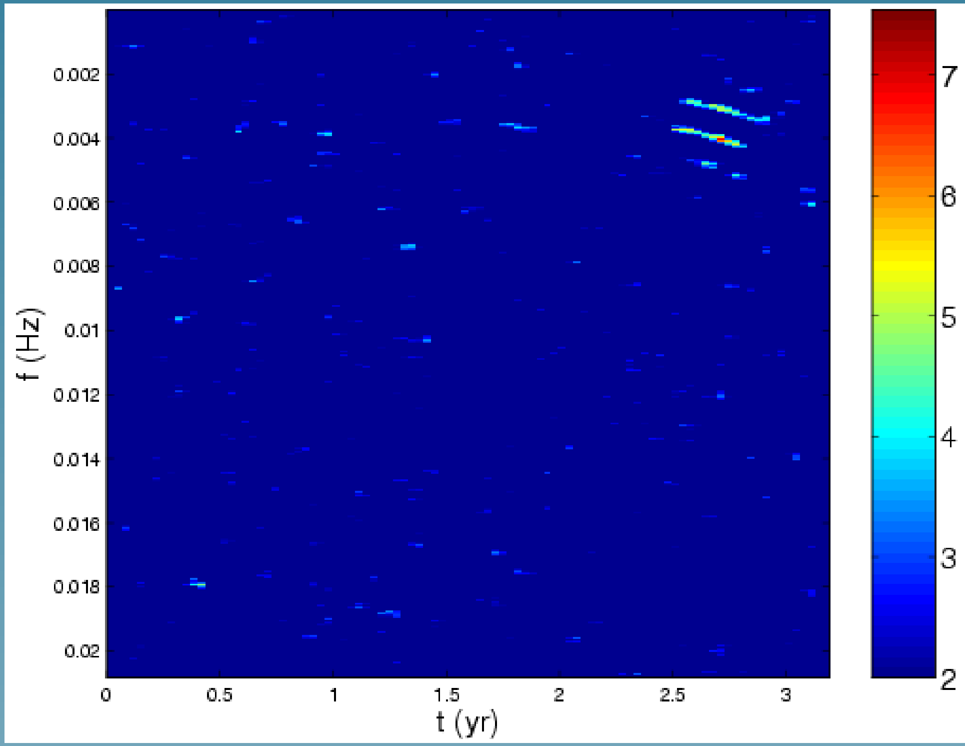
- Expect 1-10 events with $d < 1$ Gpc over three years. Also detected with high confidence.

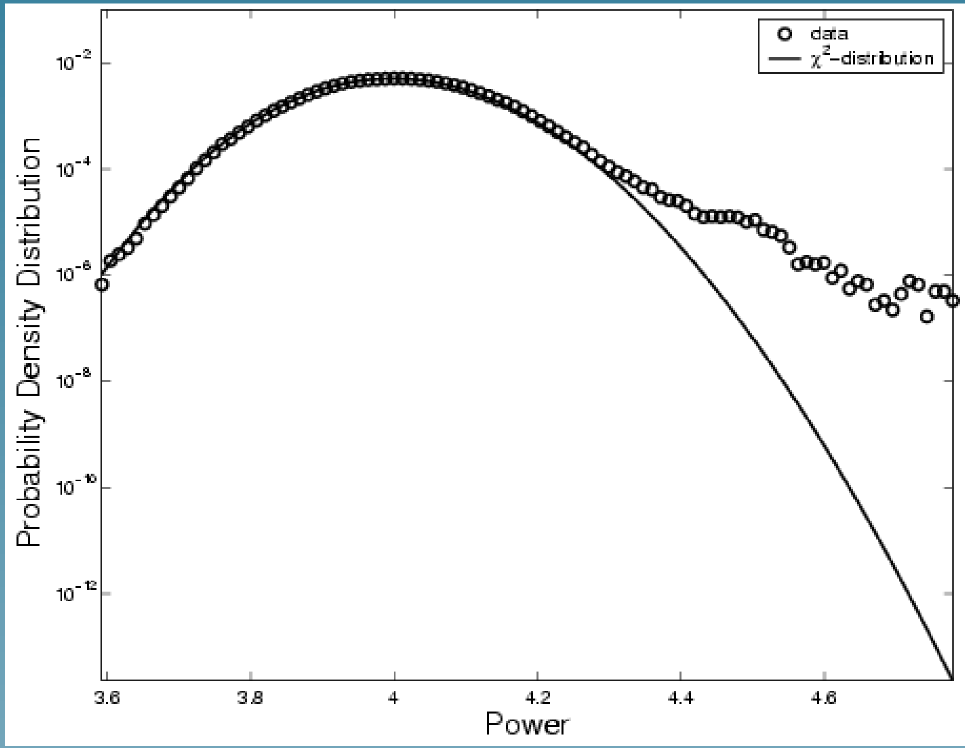




Results - source at d = 1.4 Gpc

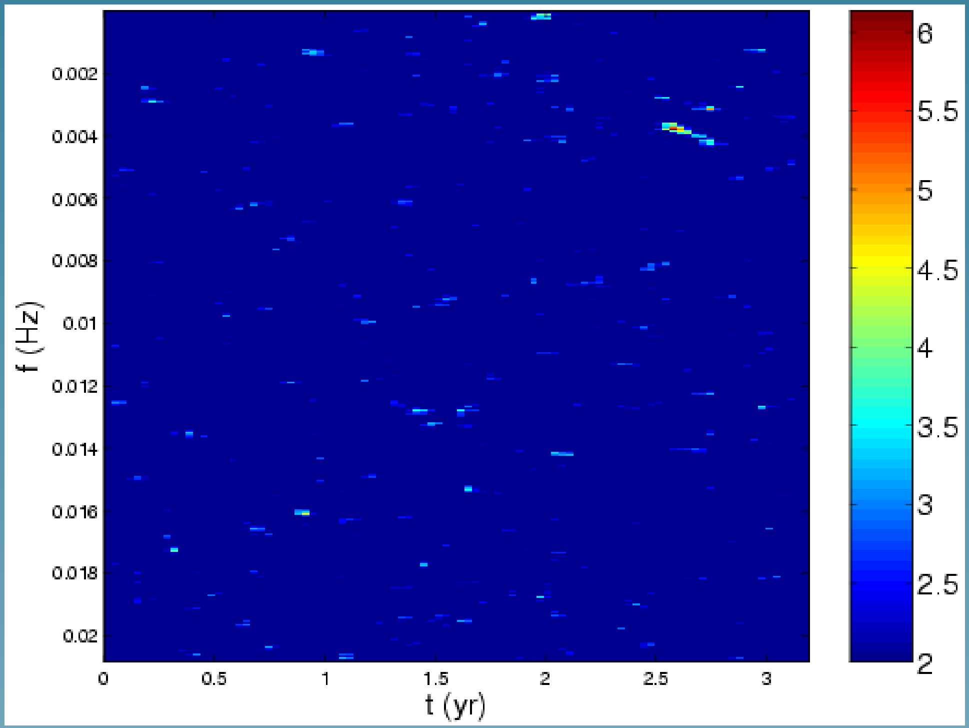
- Expect 5-40 events with $d < 1.4$ Gpc over three years. Detected with reasonable confidence.

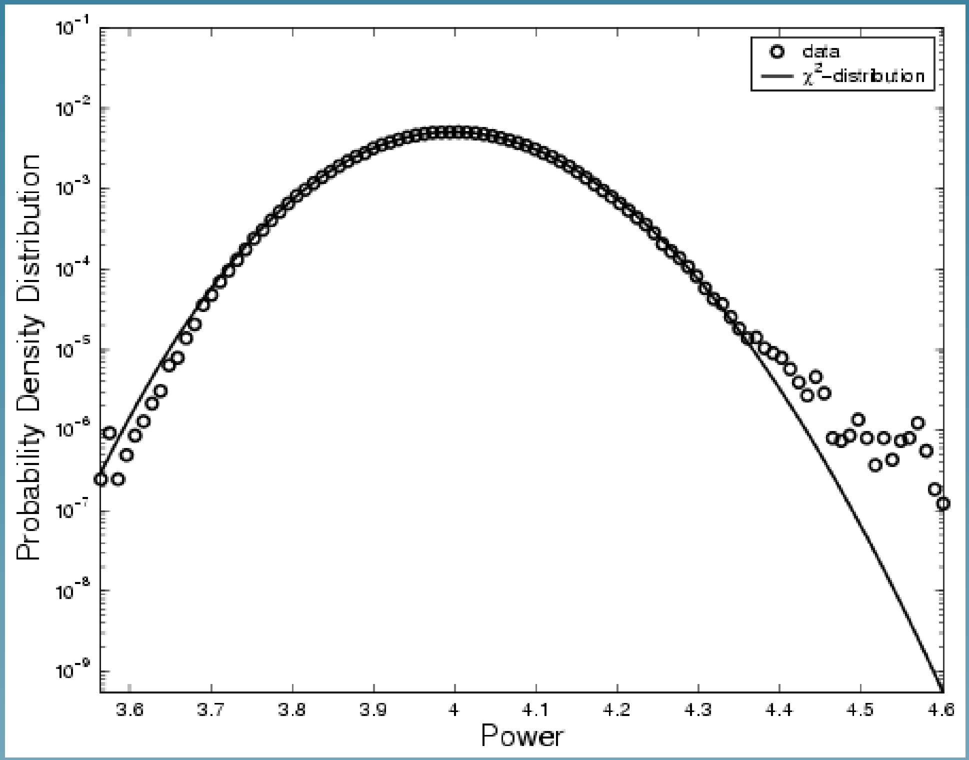




Results - source at d = 2 Gpc

- Limit of detectability. Expect 10-100 events with $d < 2$ Gpc over three years. Marginal detection.





Outstanding Issues

- Confusion is a very important issue that we have ignored so far. Confusion sources include resolvable white dwarf binaries, foregrounds from unresolvable astrophysical sources and overlap with other EMRIs.

Outstanding Issues

- Confusion is a very important issue that we have ignored so far. Confusion sources include resolvable white dwarf binaries, foregrounds from unresolvable astrophysical sources and overlap with other EMRIs.

- This approach allows event detection, but not parameter determination. However, it could provide a frequency and frequency drift rate which will significantly reduce parameter space for matched filtering follow up.

Outstanding Issues

- Confusion is a very important issue that we have ignored so far. Confusion sources include resolvable white dwarf binaries, foregrounds from unresolvable astrophysical sources and overlap with other EMRIs.

- This approach allows event detection, but not parameter determination. However, it could provide a frequency and frequency drift rate which will significantly reduce parameter space for matched filtering follow up.

- We have assumed that the noise in the t-f plane is gaussian and can be normalised according to the noise spectral density, $S_h(f)$. Non-stationarity (particularly in astrophysical foregrounds) complicates noise properties.

Outstanding Issues

- Confusion is a very important issue that we have ignored so far. Confusion sources include resolvable white dwarf binaries, foregrounds from unresolvable astrophysical sources and overlap with other EMRIs.
- This approach allows event detection, but not parameter determination. However, it could provide a frequency and frequency drift rate which will significantly reduce parameter space for matched filtering follow up.
- We have assumed that the noise in the t-f plane is gaussian and can be normalised according to the noise spectral density, $S_h(f)$. Non-stationarity (particularly in astrophysical foregrounds) complicates noise properties.
- The algorithm has not been tuned (e.g., restrict to particular box sizes, choose optimal T etc.). Must assess efficiency with other and more realistic injected waveforms.

Future developments

- Could use time frequency methods to distinguish white dwarf binaries (no change in frequency as a function of time).

Future developments

- Could use time frequency methods to distinguish white dwarf binaries (no change in frequency as a function of time).
- Time-frequency methods might also be a way to detect unmodelled sources (e.g., non black hole EMRIs).

Future developments

- Could use time frequency methods to distinguish white dwarf binaries (no change in frequency as a function of time).
- Time-frequency methods might also be a way to detect unmodelled sources (e.g., non black hole EMRIs).
- Can investigate more sophisticated search algorithms (e.g., different pixel shapes, Hough transform) that exploit the distinctive shapes of EMRI vs. binary tracks.

Future developments

- Could use time frequency methods to distinguish white dwarf binaries (no change in frequency as a function of time).
- Time-frequency methods might also be a way to detect unmodelled sources (e.g., non black hole EMRIs).
- Can investigate more sophisticated search algorithms (e.g., different pixel shapes, Hough transform) that exploit the distinctive shapes of EMRI vs. binary tracks.
- Could restrict search to bands in frequency to avoid some confusion problems. Should attempt to understand how confusion in these algorithms compares to other techniques.

Summary

- We have investigated a simple excess power technique for detecting EMRI's in the LISA data stream as a proof of principle that such methods could be used.

Summary

- We have investigated a simple excess power technique for detecting EMRI's in the LISA data stream as a proof of principle that such methods could be used.
- We find that a typical source could be detected up to a distance of ~ 2 Gpc, which would mean as many as 100 detections by this technique during the mission.

Summary

- We have investigated a simple excess power technique for detecting EMRI's in the LISA data stream as a proof of principle that such methods could be used.
- We find that a typical source could be detected up to a distance of ~ 2 Gpc, which would mean as many as 100 detections by this technique during the mission.
- Could be used as a first pass to find the loudest EMRI events, before following up with matched filtering to estimate parameters and find weaker signals.

Summary

- We have investigated a simple excess power technique for detecting EMRI's in the LISA data stream as a proof of principle that such methods could be used.
- We find that a typical source could be detected up to a distance of ~ 2 Gpc, which would mean as many as 100 detections by this technique during the mission.
- Could be used as a first pass to find the loudest EMRI events, before following up with matched filtering to estimate parameters and find weaker signals.
- Algorithm can be developed and improved in many ways and confusion issues must be carefully examined. Nonetheless, it is promising that even this simple algorithm could be useful for data analysis.