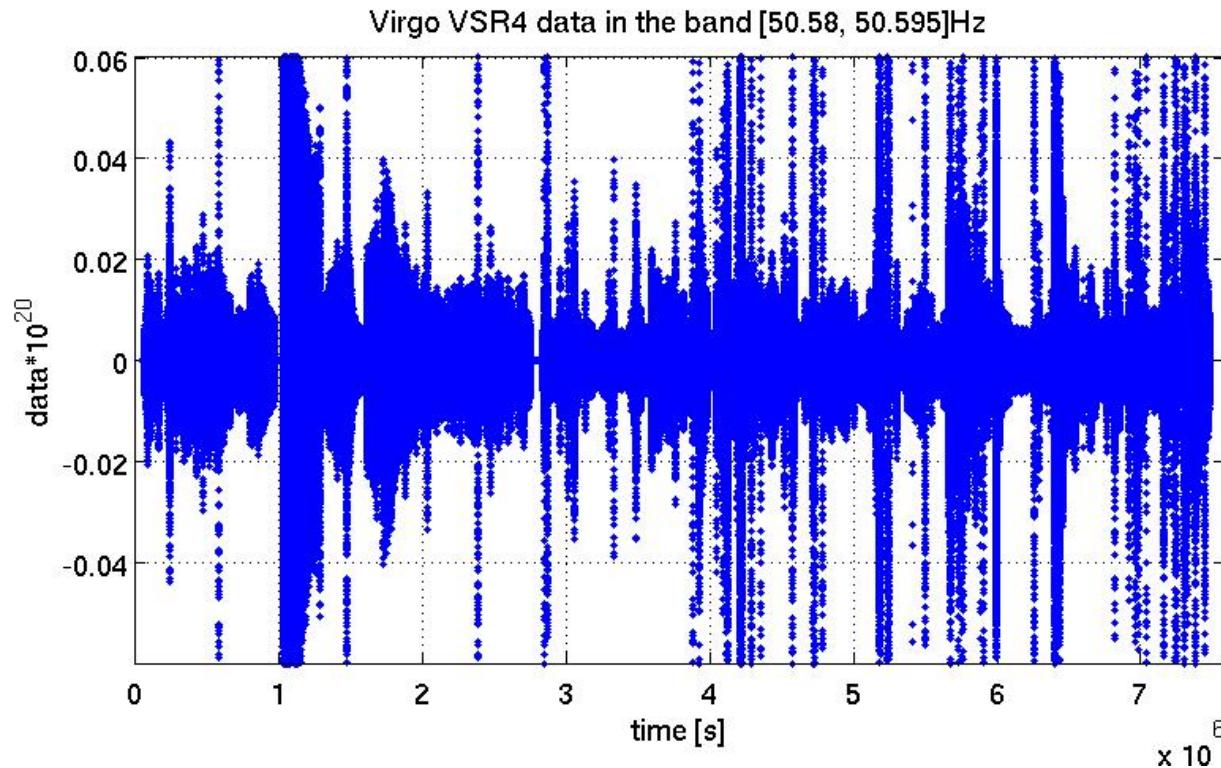


Problematiche di “analisi dati” ed estrazione del segnale dal rumore. Alcune considerazioni generali ed esempi

- Typically, we want to study the data produced by a given experimental apparatus, e.g. a particle detector, a radio-telescope, a gravitational wave detector.
- The data to be analyzed are made of instrumental or external noises, and possibly contain the signal of interest.
- **Typically, the signal is completely buried in the noise.**
- We want to:
 - Establish if a signal is present into the data
 - Estimate its characteristics and parameters

Riuscite a vedere un “segnale” qui ?

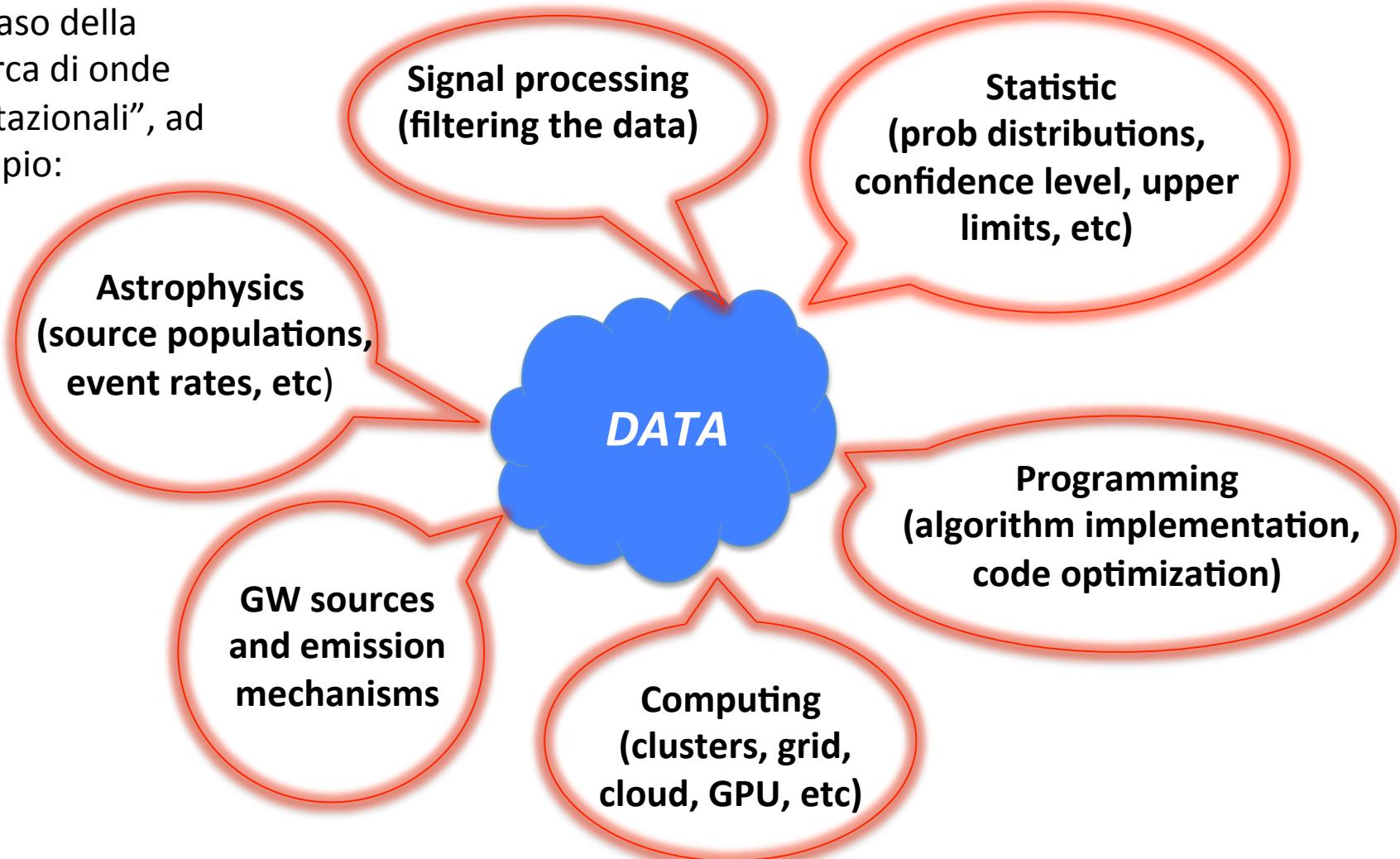
Con apposite procedure di “estrazione del segnale dal rumore” questo e’ possibile (ovviamente in funzione di un certo numero di parametri, primo fra i quali il “rapporto segnale rumore”)



Questo e’ un esempio di dati all’ uscita del rivelatore, in funzione del tempo, ossia
nel dominio del tempo

I problemi legati alla analisi dati riguardano moltissime discipline, non solo la fisica..

Nel caso della
“ricerca di onde
gravitazionali”, ad
esempio:



Molto spesso i dati vanno analizzati nel
“dominio della frequenza”, facendone analisi spettrali

- Spectral analysis is a powerful way to study physical processes by looking at the frequency-component content of the data.
- The main tool is the **Fourier Transform**:

Sufficient condition for the existence of the FT is that

$$x(t) \in L_p, 1 \leq p \leq 2$$

$$X(\omega) = \int_{-\infty}^{+\infty} x(t) e^{-j\omega t} dt$$

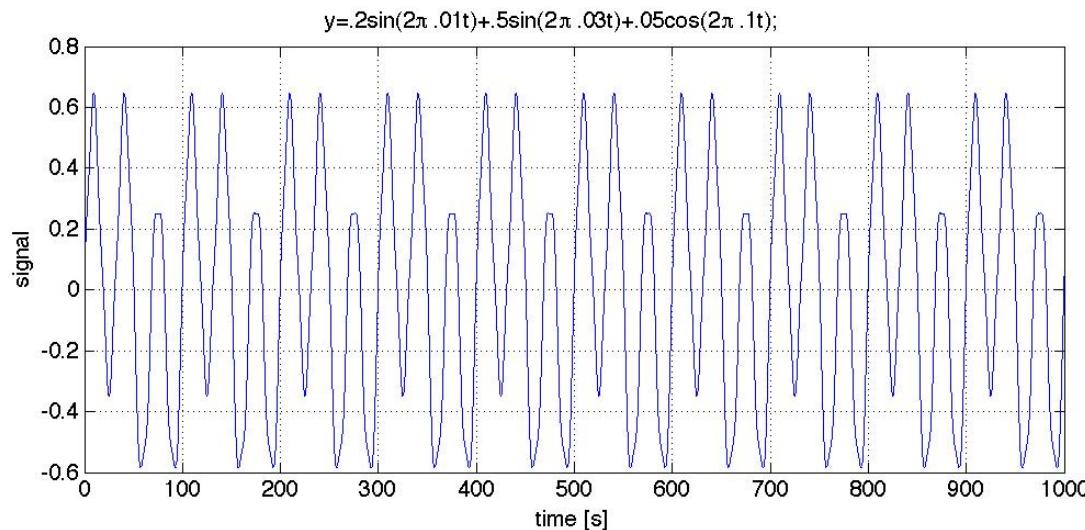
$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} X(\omega) e^{j\omega t} d\omega$$

- Parseval's theorem: $\int_{-\infty}^{+\infty} |x(t)|^2 dt = \int_{-\infty}^{+\infty} |X(\omega)|^2 d\omega$

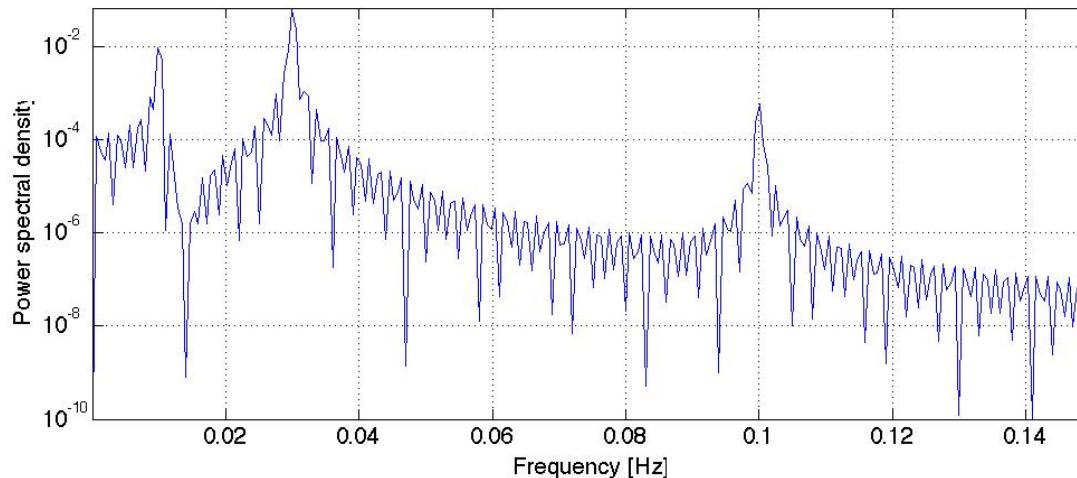
$X(\omega)d\omega$: density of signal energy in the band between ω and $\omega + d\omega$

$|X(\omega)|^2$: power spectral density, or spectral density, or energy density

If $x(t)$ is real $\rightarrow X(-\omega) = X^*(\omega)$ (hermitian) $\rightarrow |X(-\omega)|^2 = |X^*(\omega)|^2 = |X(\omega)|^2$



Dati nel dominio
del tempo.
Somma di 3
sinusoidi (solo
segnalet)



Gli stessi dati nel
dominio
della frequenza

Il segnale

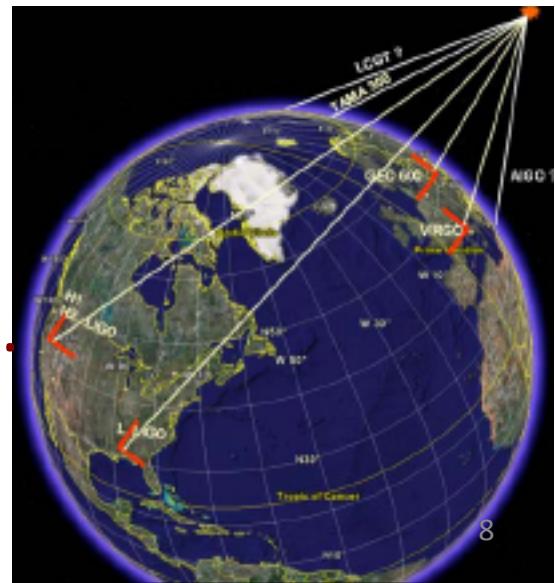
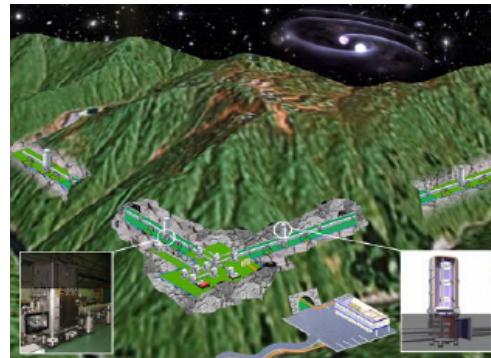
- Signal: function $y=x(t_1, t_2, \dots)$ of independent variables t_1, t_2, \dots and representing a physical quantity.
- $y_i = x(t_i), \Delta t_s = t_{i+1} - t_i$: sampling time (che va scelto seguendo regole ben precise, ad esempio rispettando il teorema del campionamento di Nyquist)
- The output of a detector is discrete and digital, as well as signals processed by a computer.
- Real or complex

Il segnale

- It can be deterministic (e.g. $y_i = \sin(t_i)$), that is represented by a mathematical formula, or random (not described by a mathematical formula).
- Random (or stochastic) signals are described by their statistical properties.
- Random signals can be studied and characterized by the theory of **stochastic processes**.



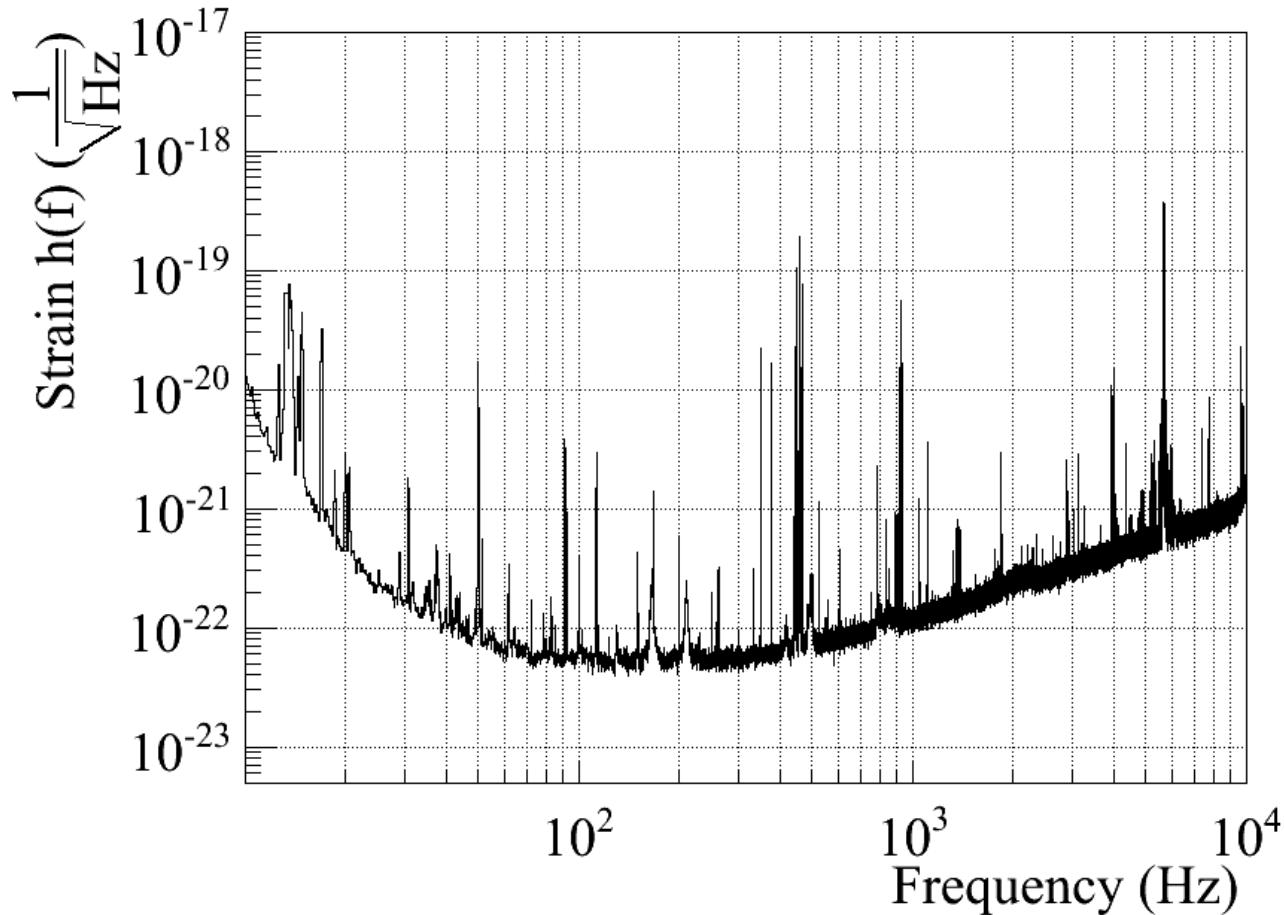
+



- A network of ITFs to improve detection probability and sky coverage.

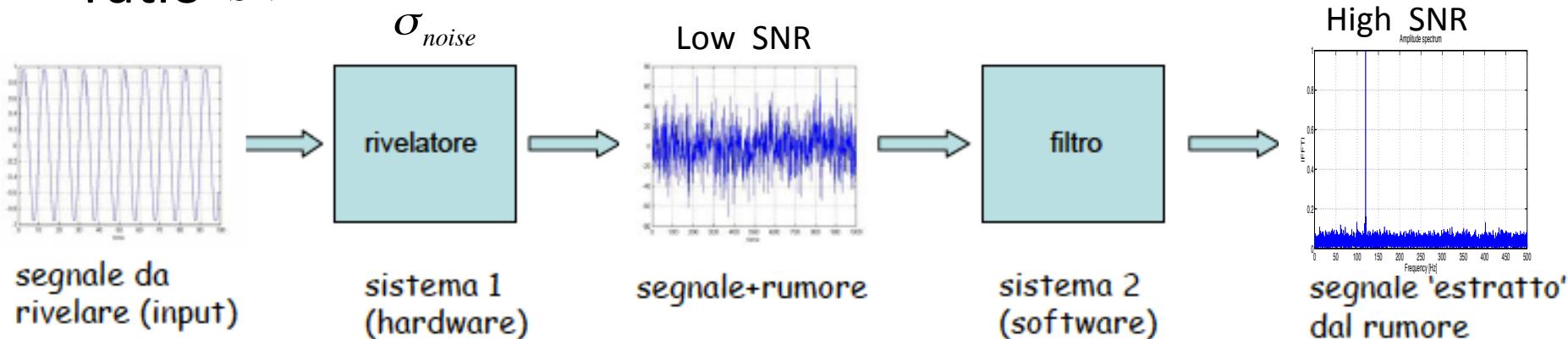
Un esempio: densita' spettrale del rivelatore di onde gravitazionali Virgo.

Virgo sensitivity. GPS: 0996606422, UTC: Fri Aug 5 19:06:47 2011



$$h(f) = \sqrt{S(f)} \quad : \text{amplitude spectral density}$$

- To analyze a set of data means to apply some operators on the data itself in such a way that the signal of interest becomes detectable and its parameters can be estimated.
- This is done by properly filtering the data. Filters are systems with particular characteristics.
- Quantitatively, we want to increase the signal-to-noise ratio $SNR = \frac{\max(x(t))}{\sigma_{noise}}$



Matched filter

- It is the linear filter that, under the assumption that noise is white Gaussian, maximizes the SNR (and also the detection probability, see later)
- It requires that the signal shape is known.
- In moltissimi casi il segnale da rivelare non e' completamente noto e questo comporta di dover implementare algoritmi basati su "banchi di filtri adattati", con conseguente aumento della potenza di calcolo necessaria.

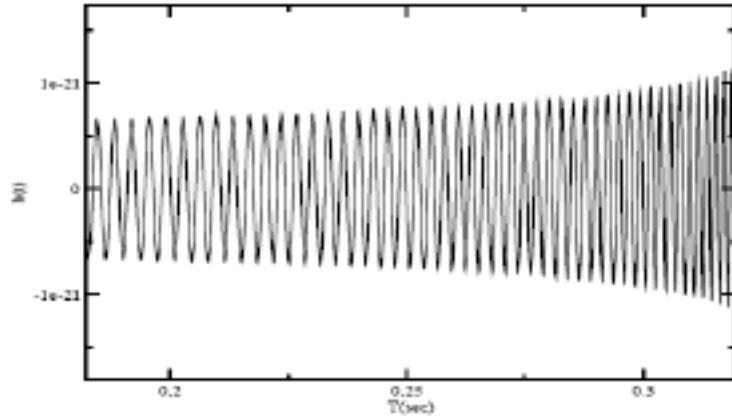
- If the signal shape is not exactly known the matched filter is not the best choice.
 - In some cases even if it could be possible to use it, in practice this is forbidden by computational reasons.
-
- **The matched filter also maximizes the detection probability at fixed false alarm probability.**

Segnali gravitazionali da binarie coalescenti

- Inspiral phase: signal with increasing amplitude and frequency:

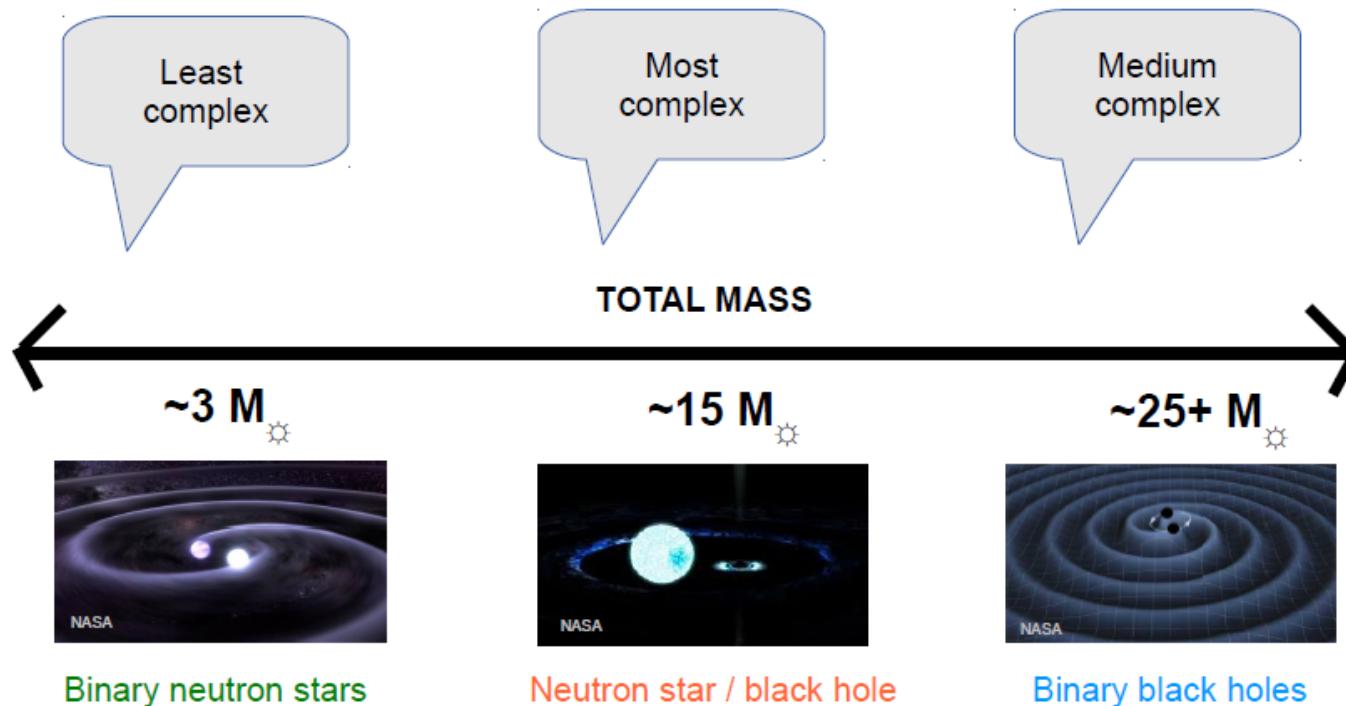
$$h \propto \frac{\mu}{r} (Mf)^{\frac{2}{3}} \cos(\Phi(f(t)) + \Phi_0)$$

$$\mu = \frac{M_1 M_2}{M} \quad \dot{f} \propto f^{11/3}$$



- The signal is accurately predictable, including also post-newtonian corrections, as long as the two objects are far from the coalescence: **matched filter can be used.**
- But: the signal shape depends on several parameters (masses, spins,...) so a bank of thousands of templates must be built and tested against the data.

Un esempio: ricerca di segnali gravitazionali da “binarie coalescenti:



Sotto certe ipotesi (durata della presa dati, numero di rivelatori) e usando come unita' di misura i "milioni di service units" (MSU), dove 1 SU= 1 core-hour su un processore di Riferimento (Xeon E5-2670):

BNS non-spinning: 1.8 MSU; **BNS spinning**: 37 MSU; **BBH**: 45 MSU; **NSBH**: 81 MSU

Una richiesta di 1.0 MSU significa che con un milione di cores finirei l' analisi in 1 ora. Per finirla ad esempio in un mese, servirebbero: $10^6/(24*30) = 1390$ cores (24 ore al giorno, per un mese)

Alcune considerazioni

- Spesso la potenza di calcolo necessaria per condurre una analisi “ottimale” dal punto di vista del risultato finale (ad esempio la rivelazione del segnale con un alto livello di confidenza) e’ “troppo elevata” (costosa, oppure tale che per raggiungere l’ energia necessaria il tempo di calcolo supera la durata della presa dati).
- Esempio nella slide seguente. *E ALLORA ?*
- *1) bisogna avere un buon modello di calcolo dell’ esperimento.*
- *2) cercare ed implementare procedure di analisi che riducano il problema del calcolo, non penalizzando troppo la sensibilita’ finale della ricerca (un esempio sono le procedure di tipo gerarchico)*
- *3) data una procedura ottimizzare gli algoritmi (anche in funzione delle caratteristiche del compilatore e delle macchine sulle quali si girera’, laddove possibile)*
- *4) Parallelizzazione e uso di risorse distribuite quali GRID (XSEDE negli US)*
- *5) altro, ad esempio uso di GPU (graphics processing units). Anche qui vanno fatto i conti con il costo del singolo processore..*

Un esempio: concentriamoci sulla prima e sulla seconda riga

1 core = 10 HSE06

Computing needs at the CCs. Preliminary estimations

Pipeline	Cores per 1 year	kHSE06	kHSE06.day
CW All-Sky $\tau_{\text{min}} = 10000 \text{ yr}$	330	3.3	1200
CW All-Sky $\tau_{\text{min}} = 1000 \text{ yr}$	33000	330	120000
CBC main analysis lhope + gwtools	300	3.0	1100
CBC Parameter Estimation and GR tests	1400-2800	14-28	5000-10000
Burst cWB offline	300	3.0	1100

1 year of data at duty cycle of 100%. Analyzed in one year.

For reference: LHC @ CNAF has used a power of 88 kHSE06

2013/05/30

15

Il CNAF (centro di calcolo nazionale dell' INFN) ha circa

13800 cores

135 kHSE06

Nota:
Algoritmo CW
gia' reso
gerarchico

therefore important to develop pipelines adaptable to different environments or interfaces which hide the different technologies to the users.

The most important issues addressed by this model may be summarized as follows:

- guarantee adequate storage and computing resources at Cascina, for commissioning, detector characterization and low-latency searches;
- guarantee fast communications between Virgo applications at Cascina and aLIGO CCs/other detectors for low-latency searches;
- guarantee reliable storage and computing resources for off-line analyses in the AdV CCs;
- push towards the use of geographically distributed resources (Grid/Cloud), whenever appropriate;
- push towards a homogeneous model for data distribution, bookkeeping and access.

Figure 1 gives a big picture of the data workflow for what concerns scientific data analysis (DA) and detector characterization (Detchar) activities in AdV.

Dal “Computing Model” del rivelatore gravitazionale Advanced Virgo

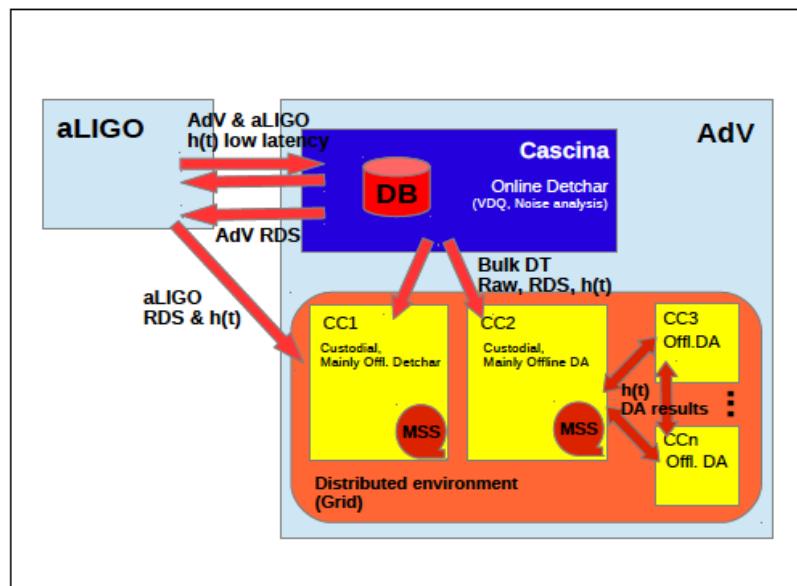


Figure 1: Data workflow for DA and Detchar activities in AdV.

Dal “Computing Model” del rivelatore gravitazionale Advanced Virgo: un modello di calcolo deve prevedere le risorse che saranno necessarie negli anni a venire. Le risorse vanno pianificate per tempo.

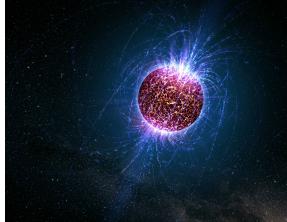
8.6 Computing needs: summary tables in regime situation (2018+) in AdV CCs

Pipeline needs in kHS06 power	local	GRID/CLOUD
Detchar Data Quality	1	–
Detchar Noise studies	1	1 ?
BURST	negl	3
CBC	–	33+
CW	–	60+
STOCHASTIC	negl.	negl.
TOTAL	2 ?	97+

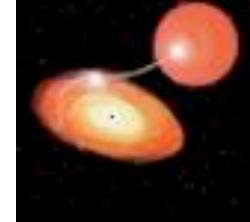
Table 8.11: Summary Table: Estimation of the computing needed locally in the CCs and under GRID/CLOUD at a regime situation (2018+), under certain hypotheses on the parameter space covered. Units are power in kHS06. The “+” indicates that this is the minimal request, with more resources we could cover a wider parameter space

Esempi utilizzando la ricerca di segnali continui (ossia sempre presenti durante la presa dati) da stelle di neutroni isolate

- 1) Algoritmi basati su “Markov Chain Monte Carlo (MCMC) technique” per rendere possibile ed accelerare, riducendo l’ energia computazionale complessiva richiesta, la stima dei parametri della(e) sorgente(i)
- 2) Algoritmi basati su procedure gerarchiche



Continuous wave signals



- CW signals are expected to be emitted by various sources containing neutron stars: tri-axial, wobbling, accreting, in binary systems,...
- **We KNOW that potential sources of CW exist:** ~2,000 NS are observed in EM (mostly pulsar), 1 billion expected to exist in the Galaxy
- **We DO NOT KNOW the amplitude of the emitted signals**

$$h_+ = h_0 \cdot \left(\frac{1 + \cos^2 \iota}{2} \right)$$

$$h_x = h_0 \cdot \cos \iota$$

$$h_0 = \frac{4\pi^2 G}{c^4} \frac{I_{zz}\varepsilon f_{GW}^2}{d}$$

$$\iota \quad : \text{angle between star spin axis and line of sight}$$
$$\varepsilon = \frac{I_{xx} - I_{yy}}{I_{zz}} \quad : \text{equatorial ellipticity}$$

$$f_{GW} = 2f_{rot}$$

The best sensitivity of current detectors corresponds to a strain of $\sim 10^{-21}$

On the other hand for a source with $f_0=100$ Hz, $\varepsilon=10^{-5}$, $d=1$ kpc we find $h_0 \sim 10^{-25}$:

$$h_0 \cong 10^{-27} \left(\frac{I_{zz}}{10^{38} \text{kg} \cdot \text{m}^2} \right) \left(\frac{10 \text{kpc}}{d} \right) \left(\frac{f}{100 \text{Hz}} \right)^2 \left(\frac{\varepsilon}{10^{-6}} \right)$$

→ signals are deeply buried into the noise!

BUT

- signal duration very long respect to typical observation times! → Signal-to-noise ratio increases with time
- signals have very specific pattern in the time-frequency plane → This helps also in rejecting noise artifacts

It is therefore important develop data analysis strategies able to detect such kind of signals and to estimate their parameters.

We distinguish two main kinds of analysis:

- ✧ Search for known neutron stars (e.g. pulsars) for which position and rotational parameters are known with high accuracy → coherent methods (like Bayes factor or “matched filter”)
- ✧ Blind searches for unknown NS → incoherent methods
(Plus intermediate cases...)

- Targeted search

If we know with high accuracy **position, frequency, frequency derivative(s)**, we can correct the Doppler effect, the spin-down etc. over long times.

In practice, we can use matched filtering over the whole observation time → coherent analysis.

Coherent analysis is the most sensitive but its computational cost rapidly increases with volume of parameter space to be explored.

$$h_{0,\min} \approx 10 \sqrt{\frac{S_n}{T_{obs}}}$$

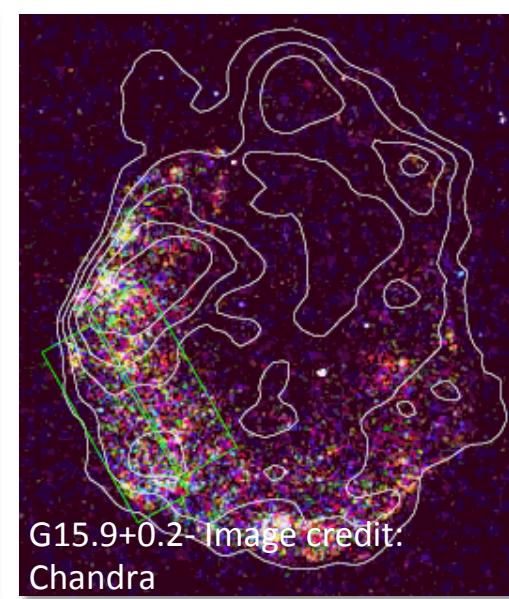
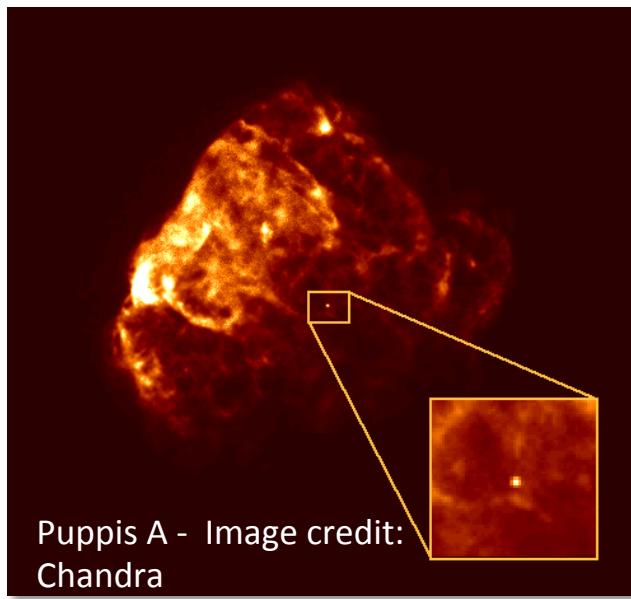
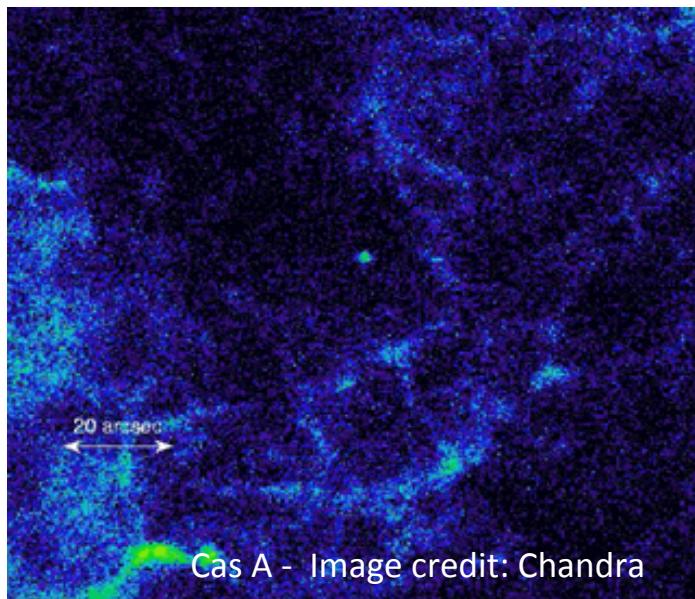
: search sensitivity

EM observations provide very accurate position and rotational parameters of many NS, especially radio pulsars.

○ Directed search

In some cases rotational parameters are not well constrained: e.g. for CCO the position is known fairly well but rotational parameters can be completely unknown (because no pulsation is observed).

→ lo spazio dei parametri incogniti aumenta



- Blind search

In blind searches we try to explore a portion of the source parameter space as large as possible:

- All-sky
- Frequency up to 1.5-2kHz
- Spin-down age $\tau = \frac{f}{\dot{f}}$ as small as possible (e.g. $< 10^3 - 10^4$ years)

This cannot be done with fully coherent methods that are computationally unfeasible because the number of points is huge ($\sim 10^{30}$).

Blind searches: Hierarchical approaches have been developed which try to satisfy two requirements:

- drastically reduce the computing power needed;
- not loose too much in sensitivity

The key idea is that of dividing data in a number of shorter segments and combine them incoherently.

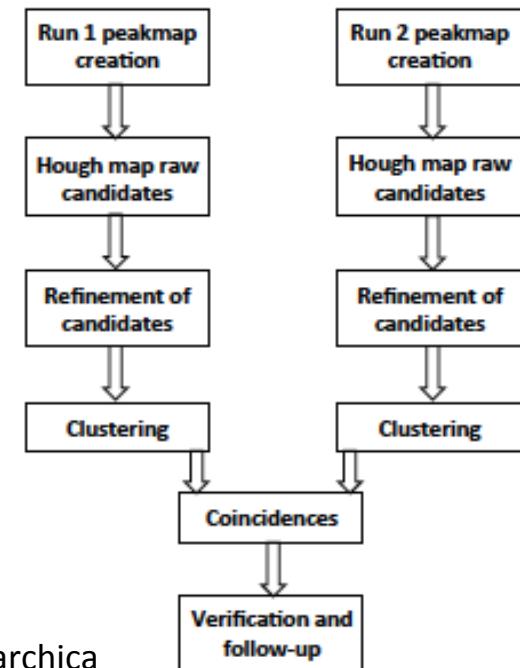
In the incoherent step a rough exploration of the parameter space is done and some candidates are selected.

Candidates are followed with a more refined search.

$$h_{0,\min} \approx \frac{10}{N^{1/4}} \sqrt{\frac{S_n}{T_{FFT}}}$$

N: number of segments

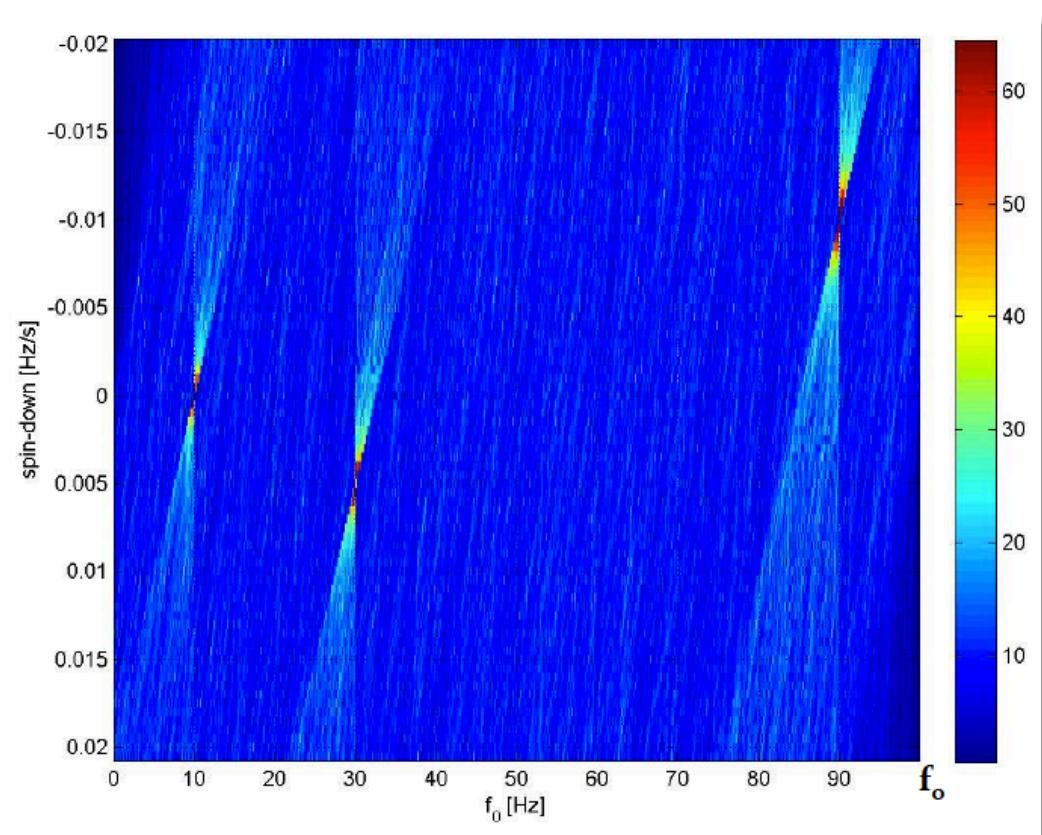
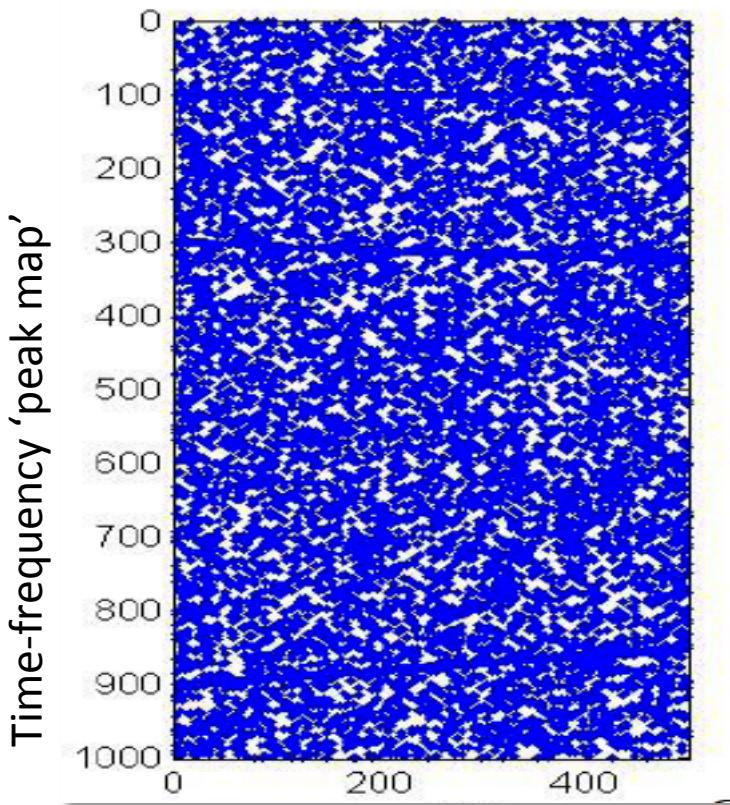
T_{FFT}: length of short pieces



Esempio di schema di procedura gerarchica

An example of incoherent step is the **Hough transform**, a pattern recognition method originally developed in the 60' to analyze tracks in bubble chambers.

It realizes a mapping between the time-frequency plane and the source parameter space.



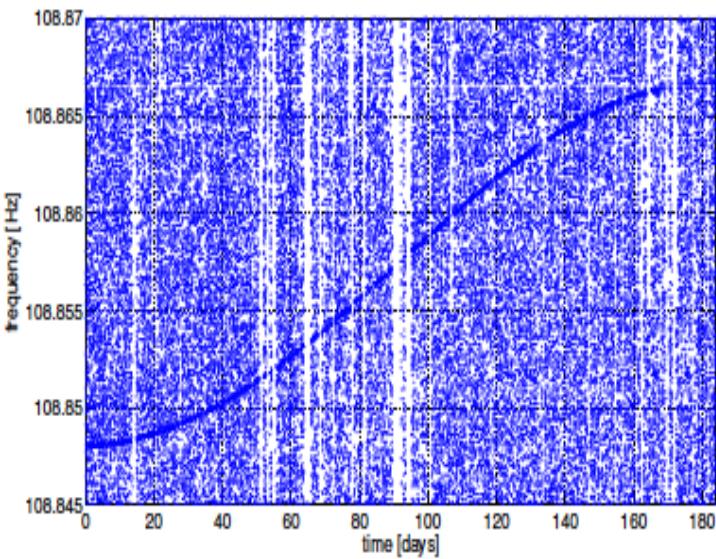
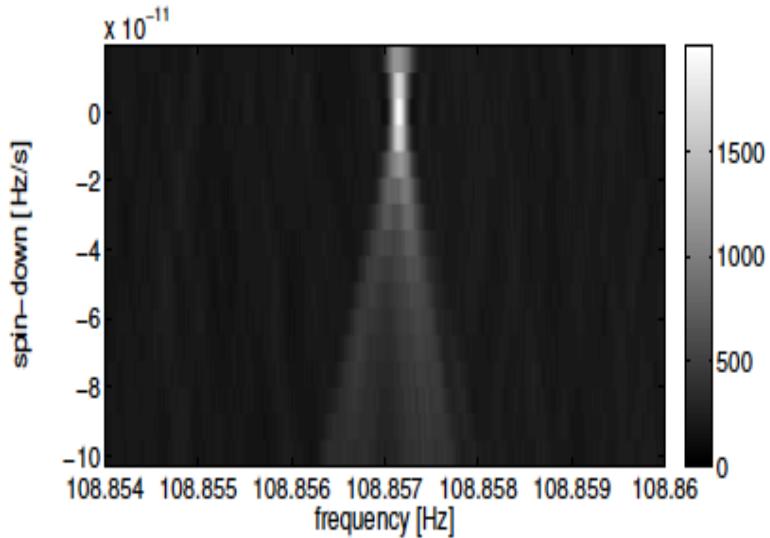


FIG. 2: Peakmap around the frequency of the HI pulsar3, with $f_0 = 108.8572$ Hz, injected in Virgo VSR2 data. Time is since the beginning of the run. The signal track is clearly visible, due to its very large amplitude.



Esempio di peakmap (sopra) e di Trasformata di Hough (sotto).

Segnale con alto SNR !

The Sky grid

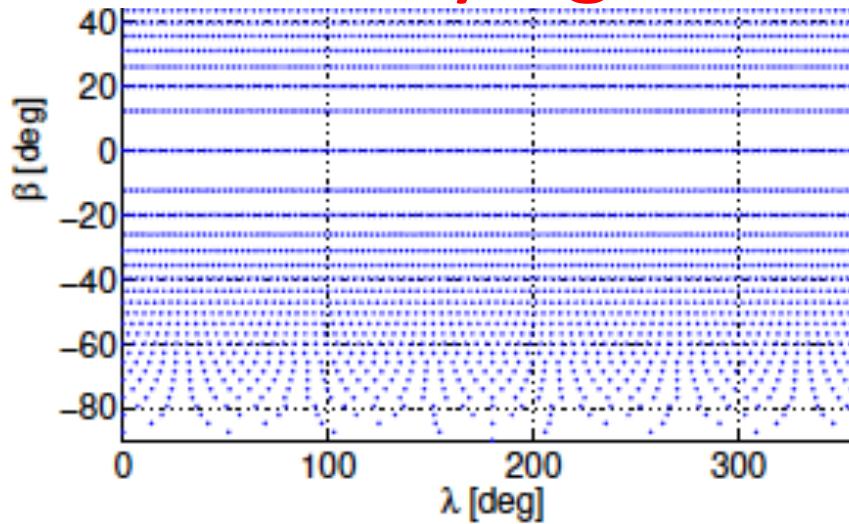
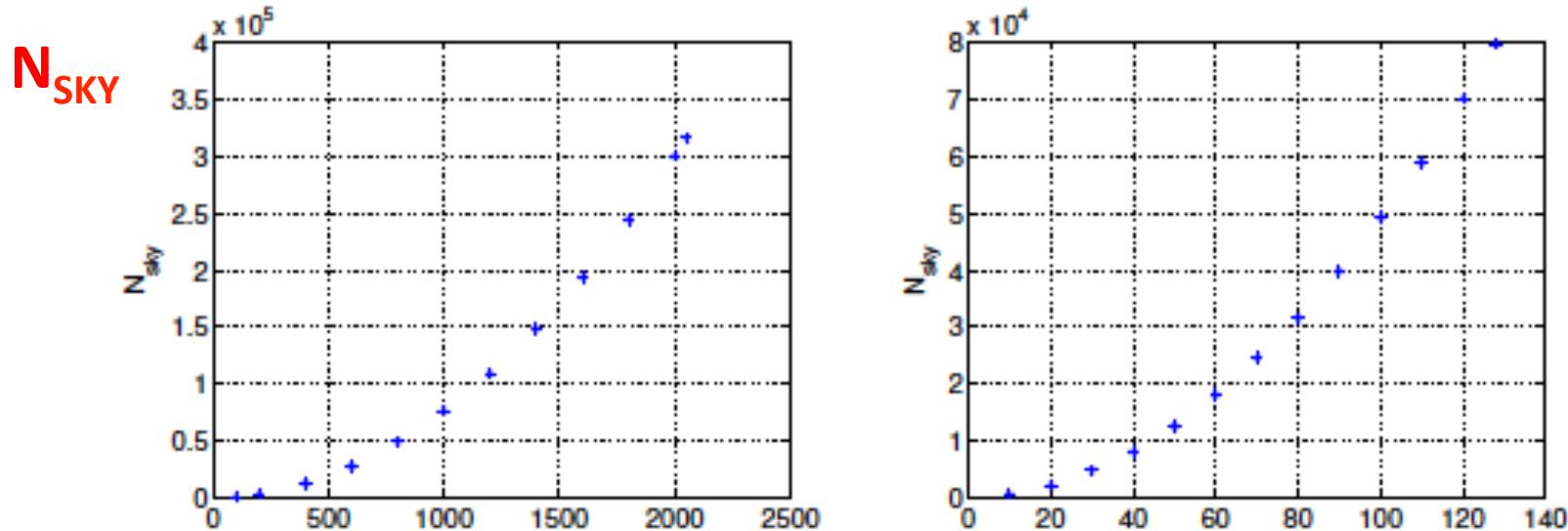
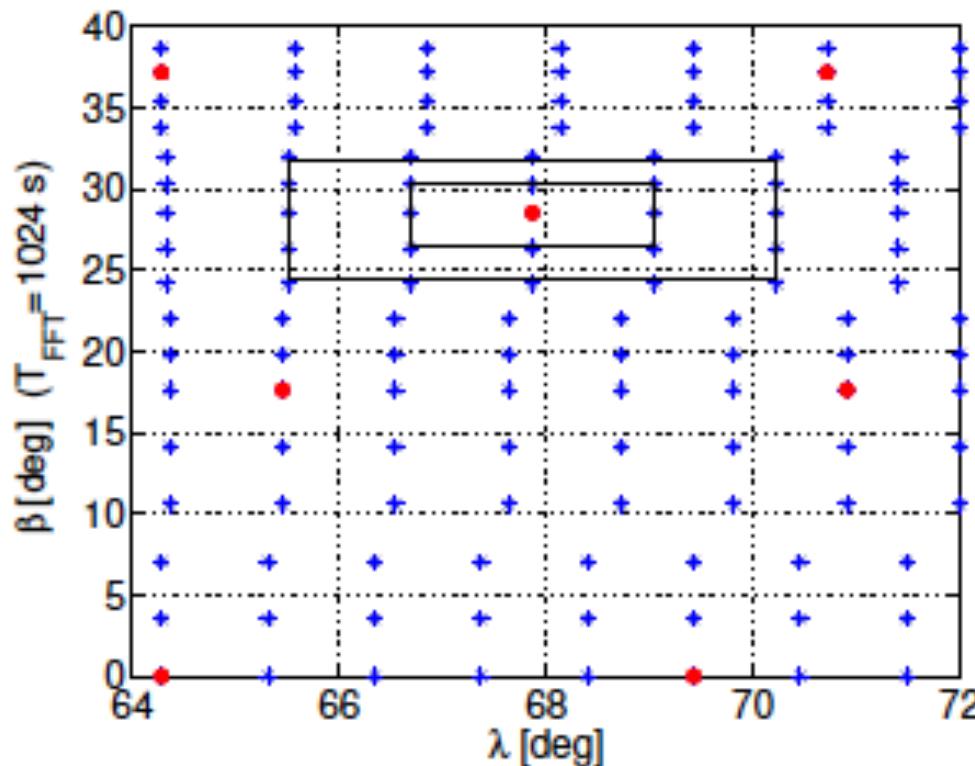


FIG. 7: Sky grid, for $T_{FPT}=1024$ s and frequency 200 Hz. $K_{sky} = 1$. x-axis: Ecliptical longitude, degrees; y-axis: Ecliptical latitude, degrees.



The refined Sky Grid



11: An example of sky grid. Red dots define points of the coarse grid, black asterisks are points of the refined grid. The rectangles defines the two "layers" that identify the refinement range around an hypothetical candidate.

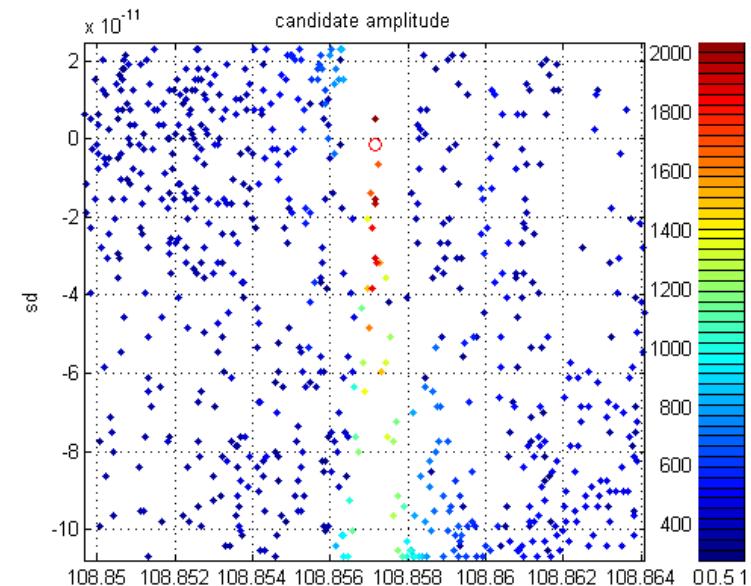
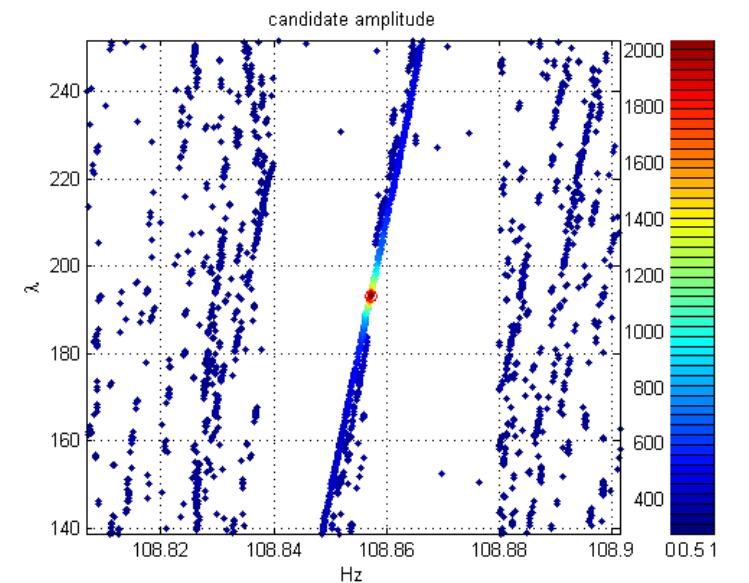
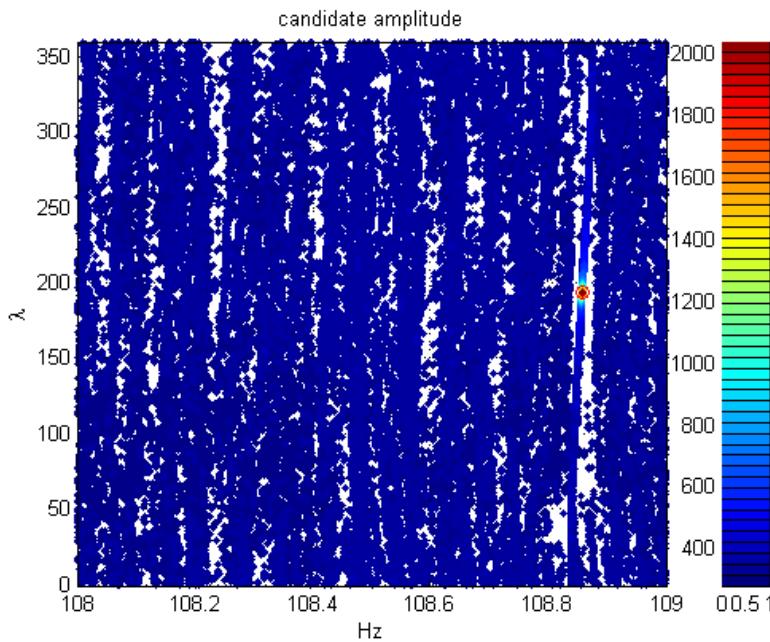
N layers = 2 mean an over-resolution factor of
Ksky = 2 N layers + 1 = 5.

Note that the over-resolution is symmetric around the coarse candidate

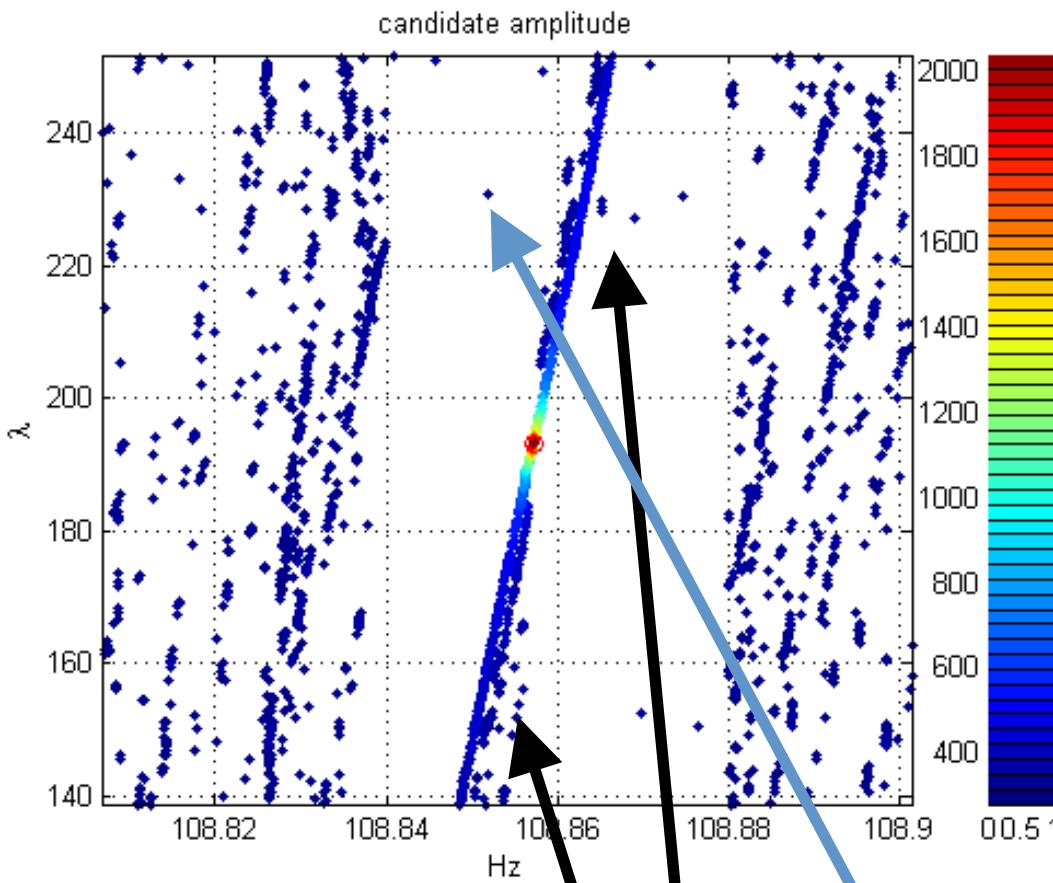
Example: pulsar 3 (removed in the analysis !)

Candidates [108-109] Hz

Here the sub-bands have been obtained
by dividing the 1 Hz band into 23 pieces



For all β and λ



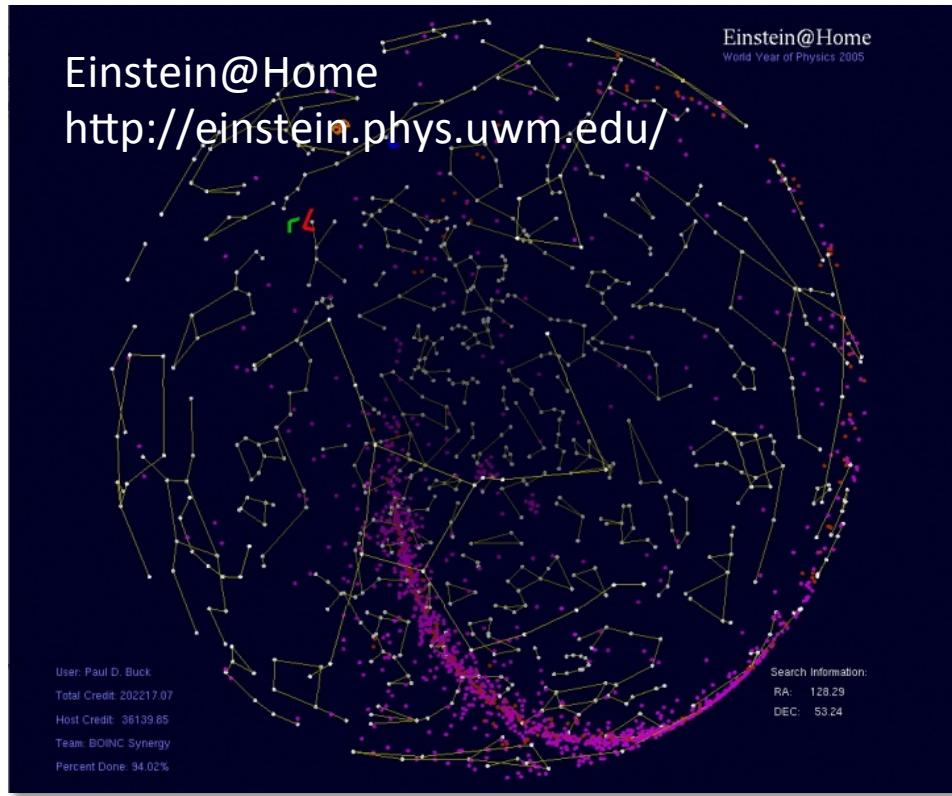
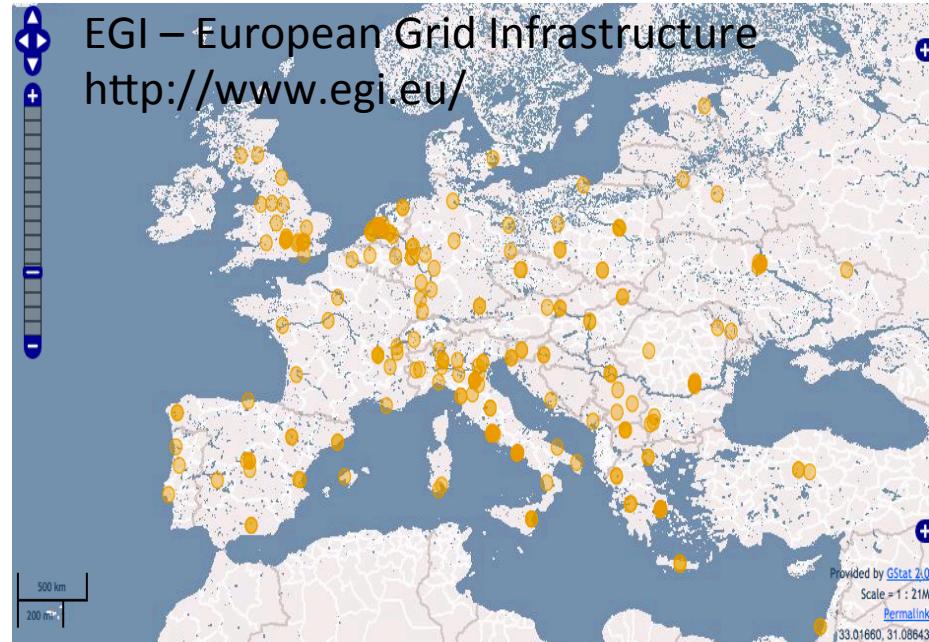
Candidates for all β s

Colours: Hough amplitude

Candidates [108-109] Hz
Around the HI (removed in the real search)
 Here the sub-bands have been obtained by dividing the 1 Hz band into 23 pieces

- The band [108-109] Hz is divided into 23 small sub-bands (23 is the number of candidates/patch at this frequency).
- Thus we select one or two candidates in each 1/23 Hz (0.043 Hz) sub-bands
- The HI is a huge signal, which implies that the second level candidates are mostly due to the HI and thus not selected, which explains the empty region around the HI
- But there are a few second order candidates selected. The one indicated by the green arrow is not due to the HI

All-sky searches are computationally-bounded: the larger is the available computing power and the deeper is the search that can be done.



In particular, going to smaller spin-down age means searching for younger, and then possibly more deformed, objects (seconda riga della tabella → la potenza di calcolo scala con il quadrato della spin-down age)

Targeted or directed

La collaborazione ha fatto una ricerca su 200 pulsar note, usando due anni di dati di due Rivelatori (LIGO Hanford, LIGO Livingston). AstroPhysical Journal, 713, 2010.

Il problema computazionale nasceva dalla stima dei parametri : la procedura consiste nel definire un modello per il segnale $y(a)$, in funzione del set di parametri incogniti, a e, usando i dati B_k , dove k rappresenta il k -esimo campione, calcolare la “posterior probability distribution” . **Statistica Bayesiana:** $p(a|d) \propto p(a) p(d|a)$ dove $d = \{B_k\}$

$$p(a|\{B_k\}) \propto \prod_j^M \left(\sum_k^n (\Re\{B_k\} - \Re\{y_k(a)\})^2 + (\Im\{B_k\} - \Im\{y_k(a)\})^2 \right)^{-m_j} \times p(a),$$

Verosimiglianza

Prior distribution

M = numero di segmenti, m_j = numero di punti nel j -esimo segmento

Fino ad un massimo di 4 parametri incogniti
(ampiezza h_0 , fase iniziale ϕ_0 , cos angolo di orientazione rispetto all' asse
del rivelatore $\cos \iota$, angolo di polarizzazione Ψ)
il problema e' gestibile con griglie (tipicamente uniformi) nello spazio dei parametri .

Targeted or directed

Se si aggiungono altri parametri , tipicamente incertezza nella frequenza e nella sua derivata, lo spin-down, l’ uso della griglia diventa computazionalmente proibitivo (o comunque troppo tempo di calcolo sarebbe necessario per la stima dei parametri).

Un buon metodo alternativo, usato nell’ analisi APJ 713 (ma anche in altre analisi e anche per stima di parametri nella ricerca di segnali transienti da binarie coalescenti) e’ l’uso di tecniche basate **su Markov Chain Monte Carlo (MCMC)**

L’ idea base e’ di esplorare lo spazio dei parametri in modo efficiente, minimizzando il tempo speso nelle aree che hanno bassa densita’ di probabilita’.

L’ integrazione MCMC esplora lo spazio dei parametri muovendosi da una posizione ad un’ altra, confrontando la densita’ di probabilita’ dei due punti e usando un algoritmo semplice per stabilire se lo step possa essere accettato o meno. Se accettato ci si muove alla nuova posizione e si ripete la procedura, se rigettato si torna alla posizione precedente e si ripete da li’.

Le iterazioni della catena vengono tutte registrate e alla fine il tempo speso in una porzione dello spazio dei parametri e’ direttamente proporzionale alla densita’ di probabilita’ a posteriori.

I nuovi punti vengono selezionati a caso da una specifica distribuzione di probabilita’ (“proposal distributions”), ad esempio una Gaussiana multivariata.

Nell’ analisi APJ 713, il guadagno in tempo di calcolo e a parita’ di efficienza e’ stato un fattore 3

Targeted or directed

- Il metodo va pero' applicato in modo da essere efficiente (deve esplorare bene lo spazio dei parametri) e bisogna minimizzare il tempo necessario alla convergenza (ossia il numero di iterazioni della catena necessario ad arrivare alla stima delle distribuzioni a posteriori per ciascun parametro, "marginalizzate" per gli altri).
- Per far questo in pratica ci sono diversi metodi, tipicamente basati su una fase di "burn-in", durante la quale le posizioni raggiunte dalla catena non sono memorizzate ma usate per "imparare".
- Spesso si usa "riparametrizzare" lo spazio dei parametri , e/o usare metodi "adattivi", nei quali si adatta la "proposal distribution" usando le informazioni ottenute a mano a mano che si procede. Notiamo che piu' la proposal distribution e' vicina alla posterior (ossia al nostro target) e piu' la convergenza e' veloce. Oppure si usa "delayed rejection", dove si sceglie il prossimo step basandosi su quello rigettato immediatamente prima.
- Un buon metodo per velocizzare la convergenza e' il "simulated annealing", usato appunto nella nostra analisi APJ.

Simulated Annealing (“ricottura simulata”)

- Il concetto di *annealing* deriva dalla scienza dei metalli, dov'è usato per descrivere il processo di eliminazione di difetti reticolari dai cristalli tramite una procedura di riscaldamento seguita da un lento raffreddamento.

A third technique for speeding convergence is *simulated annealing*. In simulated annealing, a ‘temperature’ parameter is introduced which, when large, allows the system to make transitions that would be improbable at temperature 1. The temperature is set to a large value and gradually reduced to 1. This procedure is supposed to reduce the chance that the simulation gets stuck in an unrepresentative probability island.

We assume that we wish to sample from a distribution of the form

$$P(\mathbf{x}) = \frac{e^{-E(\mathbf{x})}}{Z} \quad (30.9)$$

where $E(\mathbf{x})$ can be evaluated. In the simplest simulated annealing method, we instead sample from the distribution

$$P_T(\mathbf{x}) = \frac{1}{Z(T)} e^{-\frac{E(\mathbf{x})}{T}} \quad (30.10)$$

and decrease T gradually to 1.

Simulated annealing

$$\begin{aligned}
 p(\mathbf{a}|\mathbf{d}) &\propto p(\mathbf{a})p(\mathbf{d}|\mathbf{a}) \propto p(\mathbf{a}) \exp\left[-\frac{\chi^2(\mathbf{a})}{2}\right] \\
 &\propto \exp\left[-\frac{\chi^2(\mathbf{a}) - 2 \log [p(\mathbf{a})]}{2}\right] \\
 &\propto \exp\left[-\beta (\chi^2(\mathbf{a}) - 2 \log [p(\mathbf{a})])\right]. \tag{19}
 \end{aligned}$$

with inverse temperature β . During the burn-in period this inverse temperature can pass through values starting at a low value (thus high temperature) and ending up at $\beta = \frac{1}{2}$ which coincides with the posterior distribution. This simulated annealing technique was introduced by Metropolis et al. [12] and allows scanning of the whole parameter space by permitting larger steps. For the annealing schedule an exponential temperature curve is applied. For a certain number of iterations t_s , it starts with an inverse temperature β_0 until it reaches $\beta = \frac{1}{2}$. The inverse temperature follows the function

$$\beta(t) = \begin{cases} \beta_0 \exp\left[\frac{t}{t_s} \log\left[\frac{\beta}{\beta_0}\right]\right], & \text{if } 0 \leq t \leq t_s, \\ \frac{1}{2}, & \text{if } t > t_s, \end{cases} \tag{20}$$

depending on the current iteration t . Since the starting temperature is dependent on the data set which is influenced by the amplitude h_0 of the signal it has to be adapted accordingly.

Il parametro di temperatura β viene usato per rendere piu' piatta la distribuzione a posteriori $p(\mathbf{a} | \mathbf{d})$ durante la fase di burn-in, cosi' da esplorare piu' velocemente tutto lo spazio dei parametri. In aggiunta si devono usare tecniche per stabilire se si e' raggiunta la convergenza

As can be seen from Figs. 4 and 5, even with small signal level it is still possible to extract the most astrophysically important parameters. For this MCMC run there were a total of 10^6 iterations, with the first 3.5×10^5 as the burn-in.

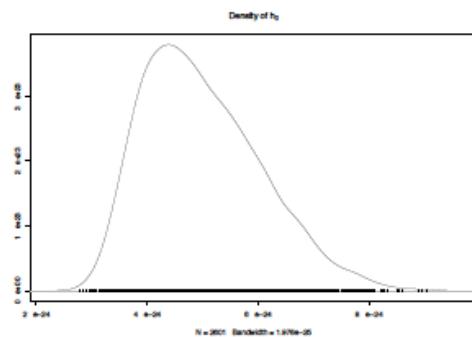


Figure 4. MCMC estimate of the posterior pdf (kernel density) for the parameter h_0 from a six parameter search using synthesized data. The real parameter value for this signal was $h_0 = 4.0 \times 10^{-24}$. This was the smallest signal detectable by the MCMC method for the noise level used.

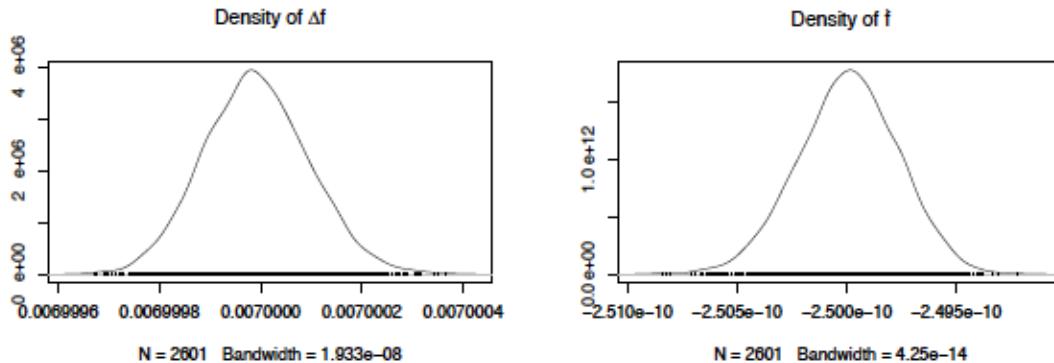


Figure 5. MCMC estimate of the posterior pdfs (kernel densities) for the parameters Δf and $\Delta \dot{f}$ from a six parameter search using synthesized data with the smallest detectable signal $h_0 = 4.0 \times 10^{-24}$. The real parameters for this signal were: $\Delta f = 7.0 \times 10^{-8}$ Hz and $\Delta \dot{f} = 2.5 \times 10^{-10}$ Hz, s $^{-1}$.

Esempio di risultati ottenuti

10^6 iterazioni

$3.5 \cdot 10^5$ per burn-in

Conclusioni

- Il mestiere del “data analyst” non e’ affatto semplice

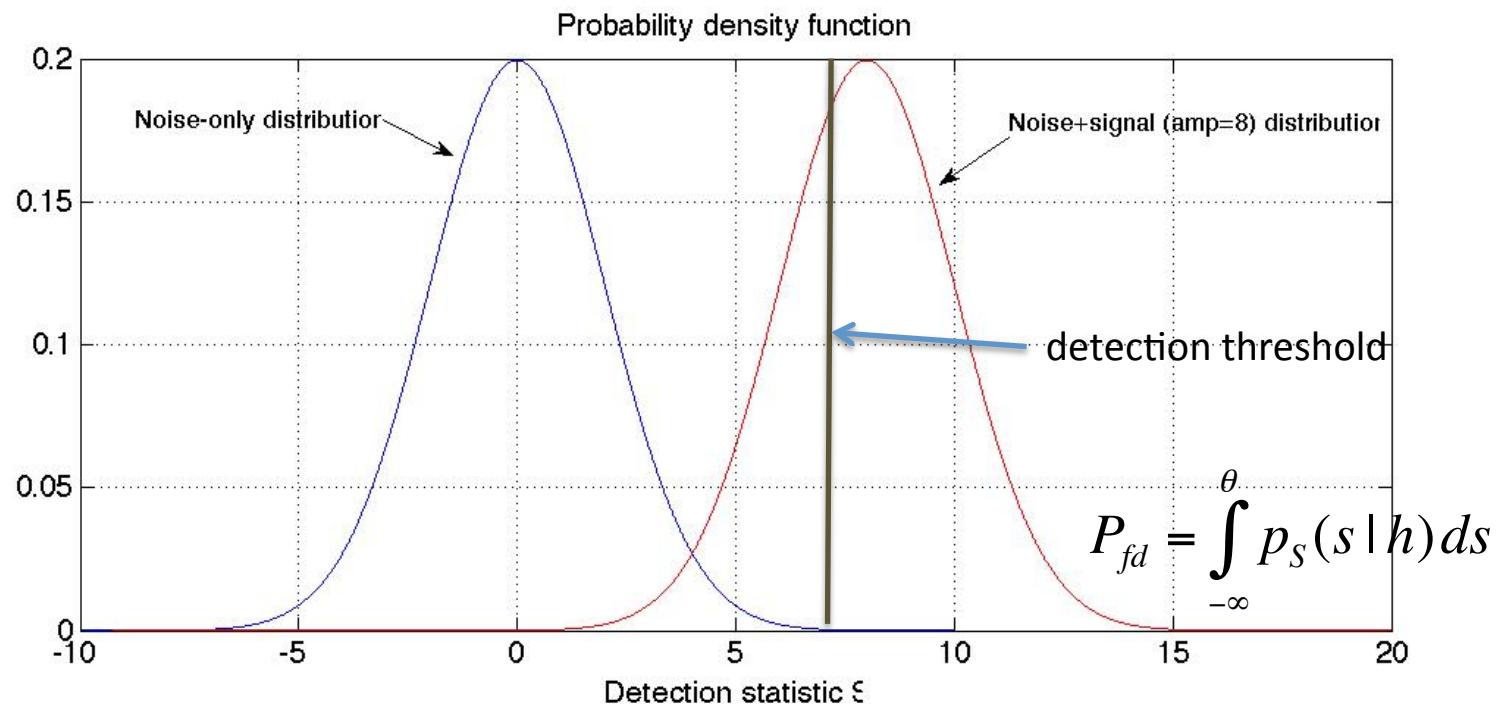
*ma e’ davvero molto interessante,
stimolante e divertente*

Detection theory

- A typical question in data analysis is: is a given interesting signal actually present into the data?
- What does it mean to detect a signal into noisy data? Qualitatively, the filter output, or a function of it, must be “large”.
- In the **frequentist** framework a suitable function of the filter output, called **detection statistic**, is computed and used to establish (statistically) if a signal is present into the data or not. **The detection statistic is a r. v.**
- The key idea is that a signal modifies the probability distribution of the detection statistic.

- In most cases claiming a detection is a statistical statement!
- There is always a, possibly small, probability that we claim a detection even if no signal is really present in the data: this is called **false alarm**.
- Whenever it is possible, we can try to improve the detection significance by e.g. analyzing more data or by using results of other experiments, etc.

- If a signal is really present in the data the probability distribution of the detection statistic changes depending on the signal amplitude.



- It may happen that even if a signal is present the value of the detection statistic is below the detection threshold: this is called a **false dismissal**.

- Equivalently, a filtering procedure can be characterized by its **sensitivity**:
the minimum signal amplitude that can be detected with a given false alarm and a given detection probability.
- The smaller is this minimum amplitude and the better it is.
- In case of detection the final analysis step consists in estimating the signal parameters.
- In case of no detection, an **upper limit** can be computed: it is the maximum signal amplitude that can be excluded to be present in the data, with a given confidence level.