Optimal detection of burst events in gravitational wave antennas

Andrea Viceré

12th October 2001

LIGO Laboratory, California Institute of Technology, MS 18-34

Abstract

In this talk I will motivate and discuss a method for the detection of burst events, which is “optimal” under the assumptions that the noise is Gaussian (and colored) and that the signal waveform is totally unknown. It is similar to the energy excess statistic proposed by Flanagan *et al.*, but differs in the choice of the prior distribution for the GW events.

Motivations

At least two astrophysical phenomena can originate short bursts of gravitational waves:

- ✔ the merger phase in a binary star (or black-hole) coalescence, and
- ✔ the collapse of the core in type-II supernovae

In both cases, it is fair to say that a large theoretical uncertainty exists on the resulting waveforms.

Models exist, but it is probably wise to be ready for the unexpected.

A zoology of GW waveforms: the Zwerger-Müller catalog

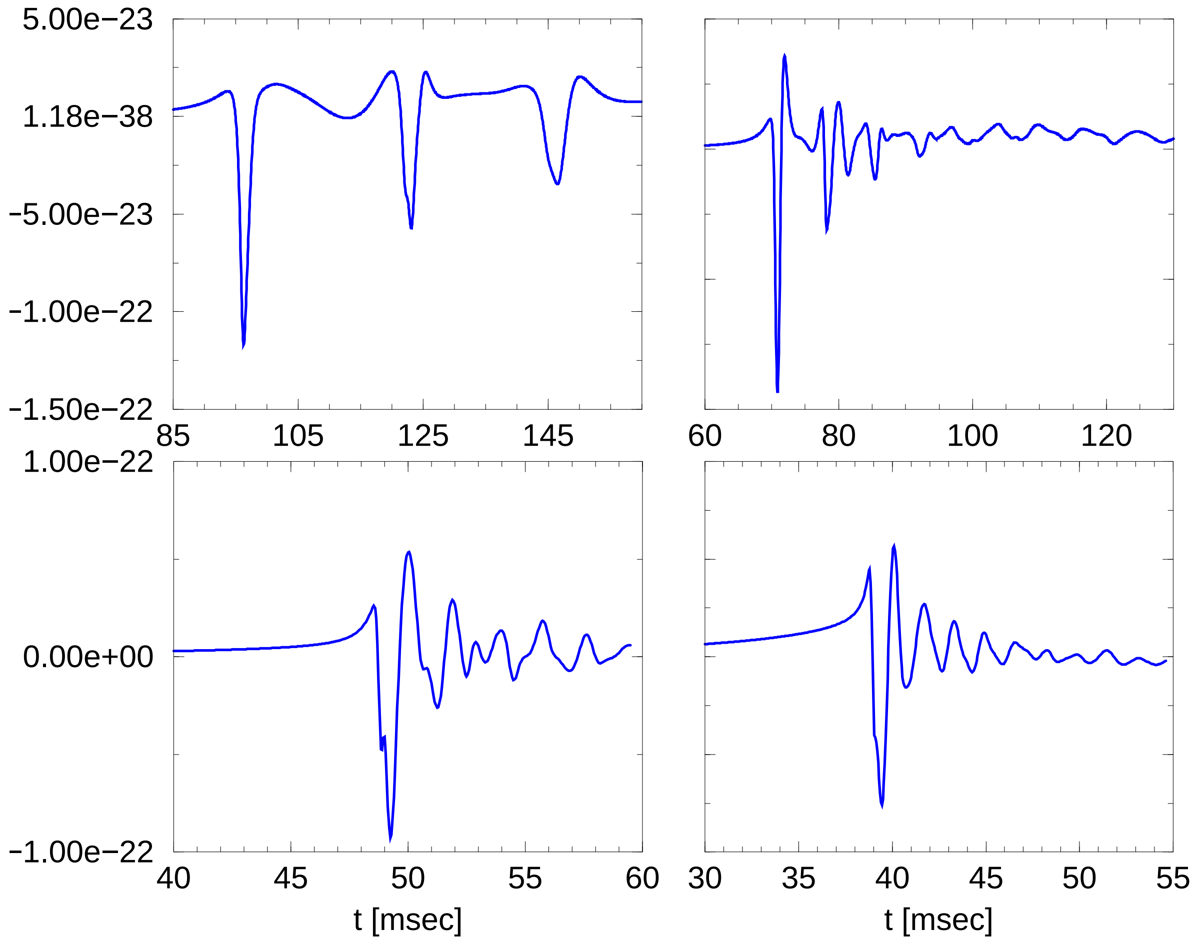
A glimpes at models helps in building some idea of what can be expected

- ✔ ZM made a simulation of the dynamics of rotational core collapse in massive stars[1].
- ✔ Axisymmetric rotating stars are assumed to be governed by a polytropic equation of state

$$P = \begin{cases} \kappa\rho^{\Gamma_1} \, \mathrm{if} \, \rho \leq \rho_{\mathrm{nuc}} \\ \kappa\rho^{\Gamma_2} \, \mathrm{otherwise} \end{cases}$$

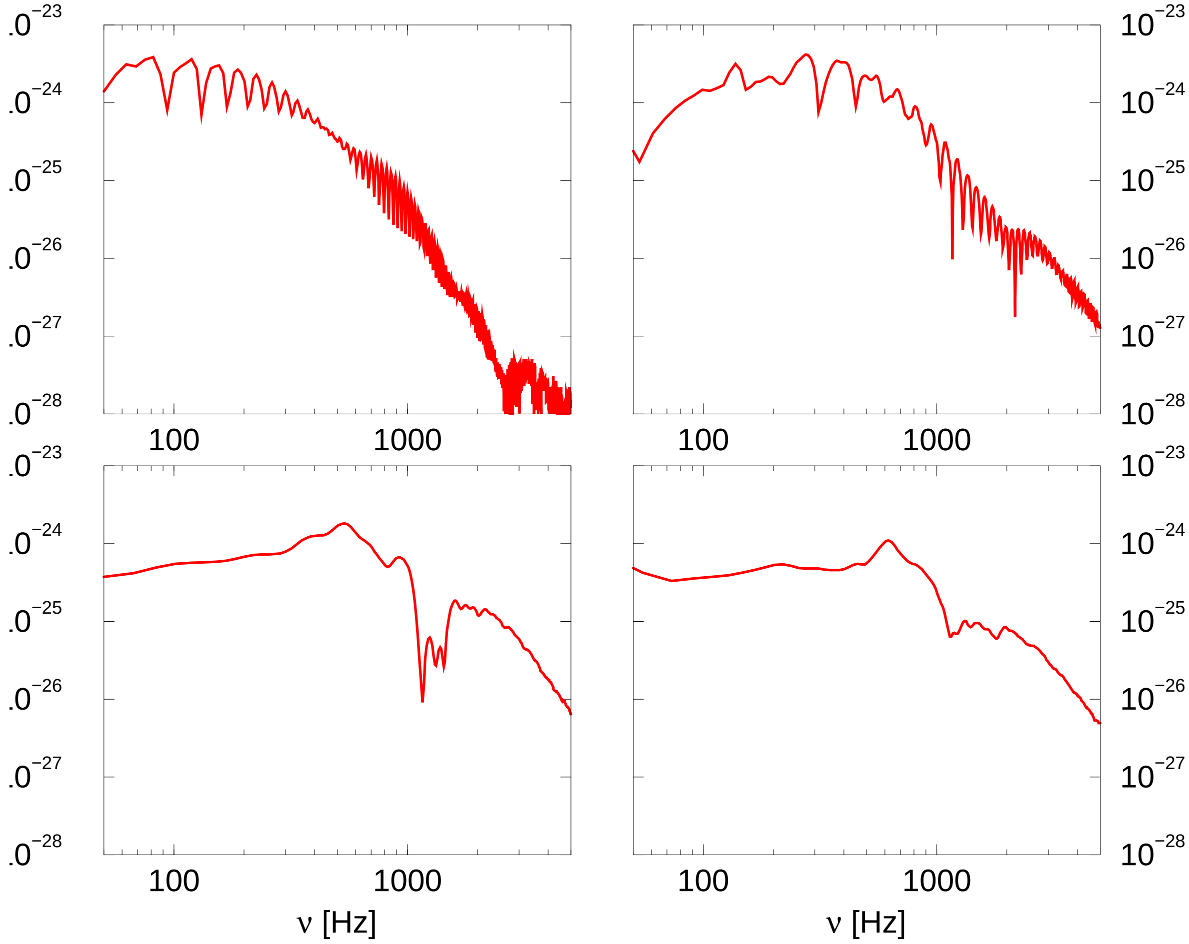
- ✔ A newtonian hydrodynamic code is used to simulate the evolution of the star during the collapse.
- ✔ Depending on the values of Γ and on the radial distribution of the angular momentum, several different waveforms result: some 78 examples are available at http://www.mpa-garching.mpg.de/ ewald/GRAV/grav.html

Some time-domain events



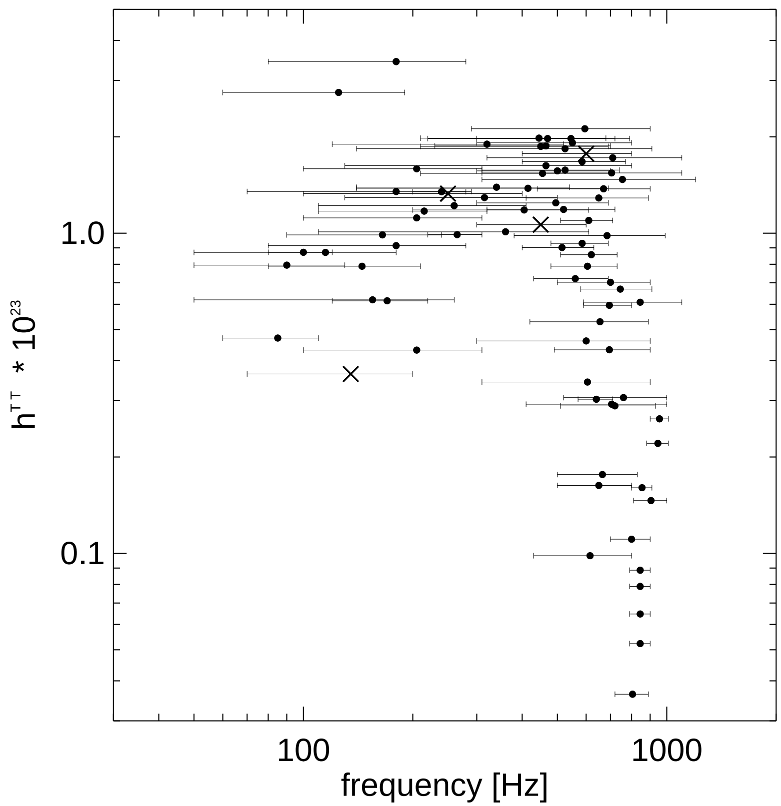
About 80 waveforms, depending on the nuclear equation of state and on the distribution of angular momentum in the star.

Spectral distribution of ZM events



Rather broadband spectra, corresponding to the relatively short events in the time domain, with features shorter than 1ms.

A look at all the ZM models together



The plot is taken from http://www.mpa-garching.mpg.de/Hydro/GRAV/grav1.html and the "error bars" indicate the frequency range containing most of the signal strength (normalized at 10Mpc).

Main characteristics of the signals

- ✔ Narrow peaks corresponding to "bounces" in the core collapse, with a 1~10 msec width.
- ✔ Oscillatory decay patterns, heavily dependent on the physical parameters.
- ✔ Roughly two signal classes: with a narrow or a broad spectral distribution of the signal.
- ✔ Typical radiated energy in the range $6 \times 10^{-11} \leq \frac{E_{\mathrm{GW}}}{M_{\odot}c^2} \leq 8 \times 10^{-8}$, and a resulting amplitude (at 10 kpc) not exceeding $O\left(10^{-20}\right)$; detectable only inside the Milky Way by LIGO or Virgo.

In short: many different waveforms, which cannot be expected to be very strong, nor are easily captured by templates: at best they can be characterized by their frequency band.

Some possible approaches to detection

✔ Set up simple algorithms which capture some of the characteristics of the waveform [see for instance Arnaud *et al.*[2, 3, 4]]

✘ δ-function filters (peak detectors) for the spikes
✘ exponentially damped sinusoids for the "after-bounce" evolution
✘ slope detectors for the (approximately) linearly growing strain, preceding the bounce

✔ Rely on methods which detect deviations from the noise background (in a given band)

✘ The *excess energy* detector by Flanagan *et al*.[5] and Anderson *et al.[6]*
✘ A *optimal* burst detector, similar to the excess energy method, whose principle and application is the topic of this talk [7].

Starting points

✔ Gaussian noise: in absence of signal, a data vector $\mathbf{n}$ is distributed as

$$P_n(\mathbf{n})d\mathbf{n} = \frac{1}{\sqrt{(2\pi)^N \det R}} \exp\left[-\frac{1}{2} n_i \left(\mathbf{R}_n^{-1}\right)_{ij} n_j\right] \prod_k dn_k;$$

with a correlation matrix $\mathbf{R}_n$.

✔ Total ignorance on the signal $\mathbf{s}$: hence

$$p_s(\mathbf{s})\, d\mathbf{s} = \prod_k ds_k$$

is *a priori* distribution which does not introduce a preferred scale.

✔ Because of the noise, the probability of observing $\mathbf{x}$ conditioned by the signal $\mathbf{s}$ is

$$P(\mathbf{x}|\mathbf{s}) = \frac{1}{\sqrt{(2\pi)^N \det R}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{s}) \cdot \mathbf{R}_n^{-1} \cdot (\mathbf{x}-\mathbf{s})\right].$$

Probability and likelyhood

Let $P(1)$ / $P(0)$ be the *a priori* probability of signal presence / absence:

$$P(\mathbf{x}) = P(\mathbf{x}|1)P(1) + P(\mathbf{x}|0)P(0)$$

is the probability of observing a data vector $\mathbf{x}$, where

$$P(\mathbf{x}|1) \equiv \int P(\mathbf{x}|\mathbf{s})p_s(\mathbf{s})d\mathbf{s}$$

is the probability of receiving $\mathbf{x}$, in presence of a signal *of any form.* We are interested in $P(1|\mathbf{x})$, the probability that some signal is present in $\mathbf{x}$: via Bayes

$$P(1|\mathbf{x}) = \frac{\Lambda(\mathbf{x})}{\Lambda(\mathbf{x}) + P(0)/P(1)}$$

where the likelihood $\Lambda(\mathbf{x})$ is

$$\Lambda(\mathbf{x}) \equiv \frac{P(\mathbf{x}|1)}{P(\mathbf{x}|0)} = \int e^{-\frac{1}{2}\mathbf{s}\cdot\mathbf{R}_n^{-1}\cdot\mathbf{s}+\mathbf{s}\cdot\mathbf{R}_n^{-1}\cdot\mathbf{x}} p_s(\mathbf{s})d\mathbf{s}$$

The choice of the prior

The choice of $p_s(\mathbf{s})$ is crucial in determining the statistic: for instance, if

$$p_s(\mathbf{s}) = f\left(\mathbf{s} \cdot \mathbf{R}_n^{-1} \cdot \mathbf{s}\right)$$

one can obtain the *excess energy* detector of Anderson *et al.* [6].

Our choice is instead to assume that the signal affects just $N_{\mathrm{burst}} \ll N$ samples, where N is the size of the data vector

and that each signal bin can assume *a priori* any value.

We search for signals in the space $\mathcal{V}_{\parallel}$, a subspace of $\mathcal{V}$; in turn a vector long enough to resolve the spectral features.

The analysis consists of sliding the analysis window, computing an appropriate statistic on each possible $\mathcal{V}_{\parallel}$ space, and applying some statistical test.

The “optimal” statistic

Integrating over all possible signals $\mathbf{s}$ with the condition that they affect just N_{burst} samples we obtain for the log-likelihood

$$\ln \Lambda(\mathbf{x}) \equiv L(\mathbf{x}) = \sum_{i,j \in \mathcal{V}_{\parallel}} \left(\mathbf{R}_n^{-1} \cdot \mathbf{x}\right)_i \left(\left(\left(\mathbf{R}_n^{-1}\right)_{\parallel} \right)^{-1} \right)_{ij} \left(\mathbf{R}_n^{-1} \cdot \mathbf{x}\right)_j .$$

where we have defined the statistic L, which needs to be estimated on each block of N_{burst} data, out of N input data.

There are essentially two analysis steps:

✔ filtering the data to obtain the vector $\mathbf{y} = \mathbf{R}_n^{-1} \cdot \mathbf{x}$;

✔ restricting the vector $\mathbf{y}$ on the $\mathcal{V}_{\parallel}$ subspace and build the scalar product $\mathbf{y} \cdot \mathbf{M}^{-1} \cdot \mathbf{y}$ with $\mathbf{M} \equiv \left(\mathbf{R}_n^{-1}\right)_{\parallel}$: the inverse of $\mathbf{R}_n$, restricted to $\mathcal{V}_{\parallel}$.

Complicated? Not really!

Filtering step

The first step

$$\mathbf{y} = \mathbf{R}_n^{-1} \cdot \mathbf{x}:$$

becomes in the Fourier domain

$$y[l] = \frac{2}{N} \sum_{k=0}^{N-1} \frac{\tilde{x}[k] \times 1}{S_n[k]} e^{-i2\pi k l/N}$$

which means Wiener filtering with δ-function template (whose DFT is a constant).

One understands that N needs to be large enough to resolve the spectral features in the input data: otherwise, the relation

$$(\mathbf{R}_n)_{ab} = \frac{1}{2} \sum_{k=0}^{N-1} S_n[k] \, e^{-i2\pi k(a-b)/N}$$

between the spectrum S_n and the correlation matrix $\mathbf{R}_n$ would not be accurate.

Assembling the statistic ...

The next step seems cumbersome: one should

- ✔ estimate the $N \times N$ matrix $\mathbf{R}_n$ and invert it, obtaining $\mathbf{R}_n^{-1}$;
- ✔ restrict the resulting matrix to indices in the $\mathcal{V}_\parallel$ subspace, obtaining a $N_{\rm burst} \times N_{\rm burst}$ matrix $\left(\mathbf{R}_n^{-1}\right)_\parallel$
- ✔ Invert again the matrix, and sandwich it between the $\mathbf{y}_\parallel$ vector, obtaining

$$L = \mathbf{y}_\parallel \cdot \left[\left(\mathbf{R}_n^{-1}\right)_\parallel\right]^{-1} \cdot \mathbf{y}_\parallel$$

This sequence of inversion, projection and inversion looks awkward, but becomes simple if we recall how $\mathbf{y}$ was obtained.

A simpler form

Recalling the definition of $\mathbf{y}$

$$\mathbf{y} \equiv \mathbf{R}_n^{-1} \cdot \mathbf{x},$$

the correlation matrix $\mathbf{R}_y$ of the process y (in absence of signal) is just

$$\mathbf{R}_y = \mathbf{R}_n^{-1};$$

hence our statistic becomes simply

$$L = \mathbf{y}_{\|} \cdot \left[\left(\mathbf{R}_y \right)_{\|} \right]^{-1} \cdot \mathbf{y}_{\|}$$

which is quite intuitive: data $\mathbf{y}$ need to be weighted with the inverse of their noise matrix $\mathbf{R}_y$.

What we need to address now is how to perform this inversion in a computationally efficient way.

A limiting case

Suppose that N_{burst} is large enough: in this case the correlation matrix can be approximated by the DFT of the corresponding noise spectrum, and

$$L \approx \frac{f_s}{N_{\mathrm{burst}}} \sum_{k=1}^{N_{\mathrm{burst}}/2-1} \frac{\left|\tilde{y}_{\|}[k]\right|^2}{S_y[k]}$$

which is similar (apart the δ-filtering step) to the *excess energy* of Anderson *et al*.

It should be understood that S_y is the spectrum at a frequency resolution f_s/N_{burst}, hence different from $S_n[k]^{-1}$ used in the δ-filtering step.

The question is however: how L can be computed if N_{burst} is not *large*?

And further, what does it mean *large* or *small* in this context?

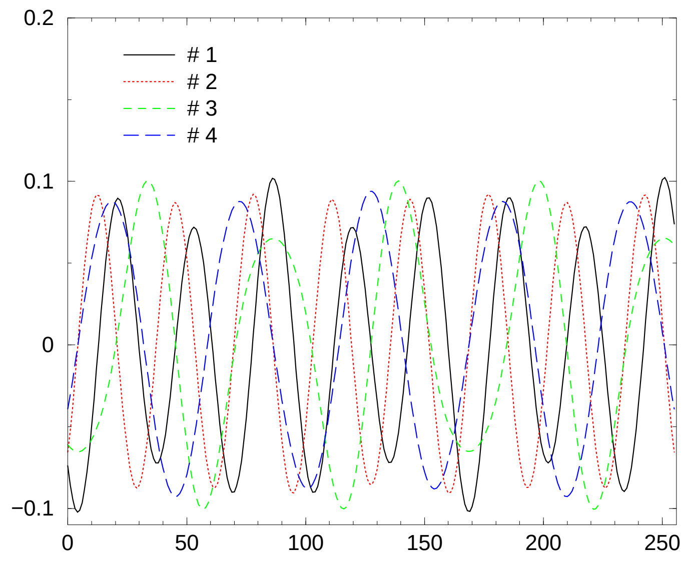
To answer both question we need to write a convenient exact form of L.

The Karhunen-Loève decomposition

The matrix $(\mathbf{R}_y)_{\|}$ has eigenvectors ψ^k and eigenvalues σ_k: therefore

$$(\mathbf{R}_y)_{\|} = \sum_{k=1}^{N_{\mathrm{burst}}} \sigma_k \psi^k \otimes \psi^k \quad \text{and} \quad (\mathbf{R}_y)_{\|}^{-1} = \sum_{k=1}^{N_{\mathrm{burst}}} \frac{1}{\sigma_k} \psi^k \otimes \psi^k$$

the eigenvectors look like sines and cosines



here are examples of vectors appropriate for LIGO 40m data.

What is the advantage of the DKL?

Given the DKL decomposition

$$\mathbf{y}_{\parallel} = \sum_{k=1}^{N_{\mathrm{burst}}} c_k \psi^k$$

the burst statistic is simply

$$L = \sum_{k=1}^{N_{\mathrm{burst}}} \frac{c_k^2}{\sigma_k};$$

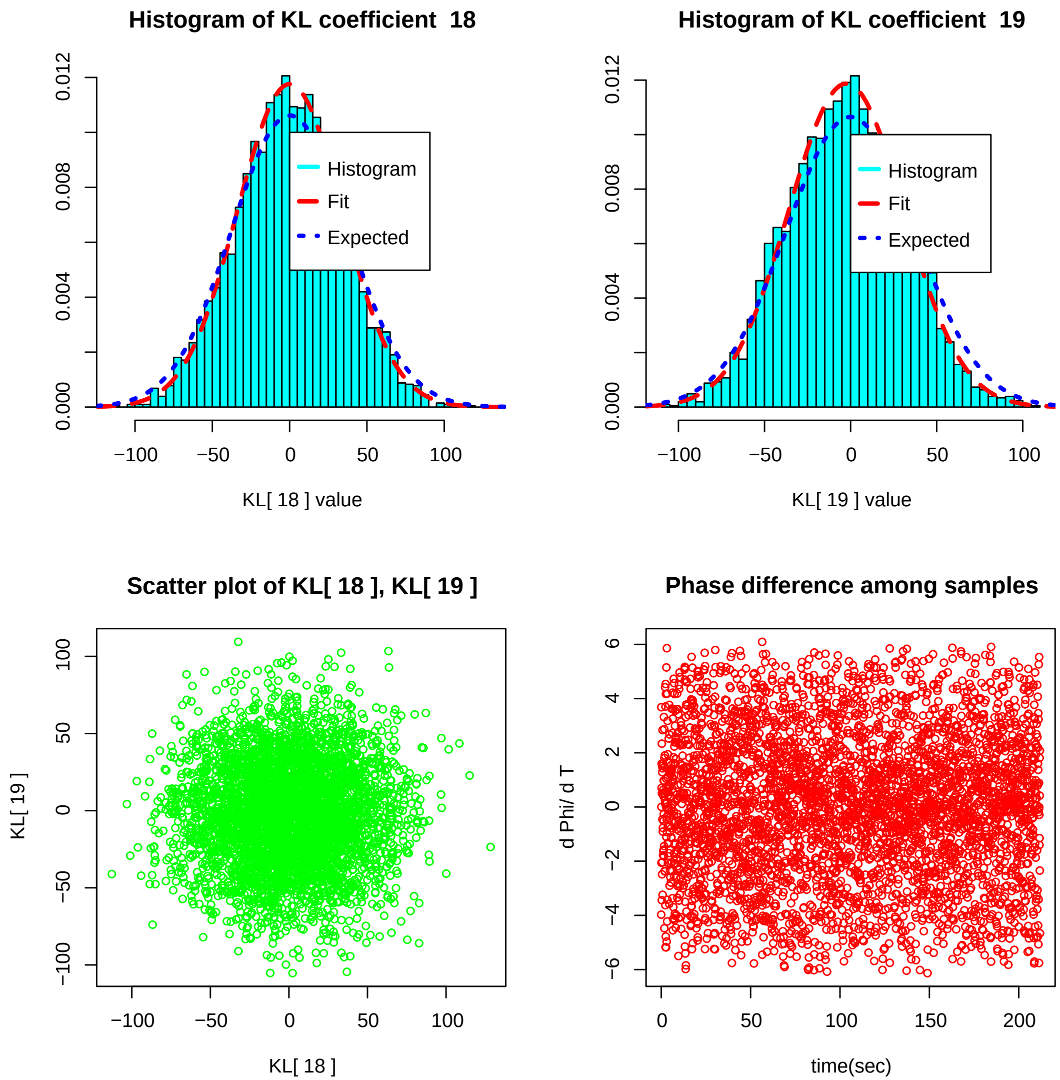
the c_k coefficients are statistically uncorrelated

$$E\left[c_k c_l\right] = \sigma_k \delta_{kl}$$

and Gaussian distributed.

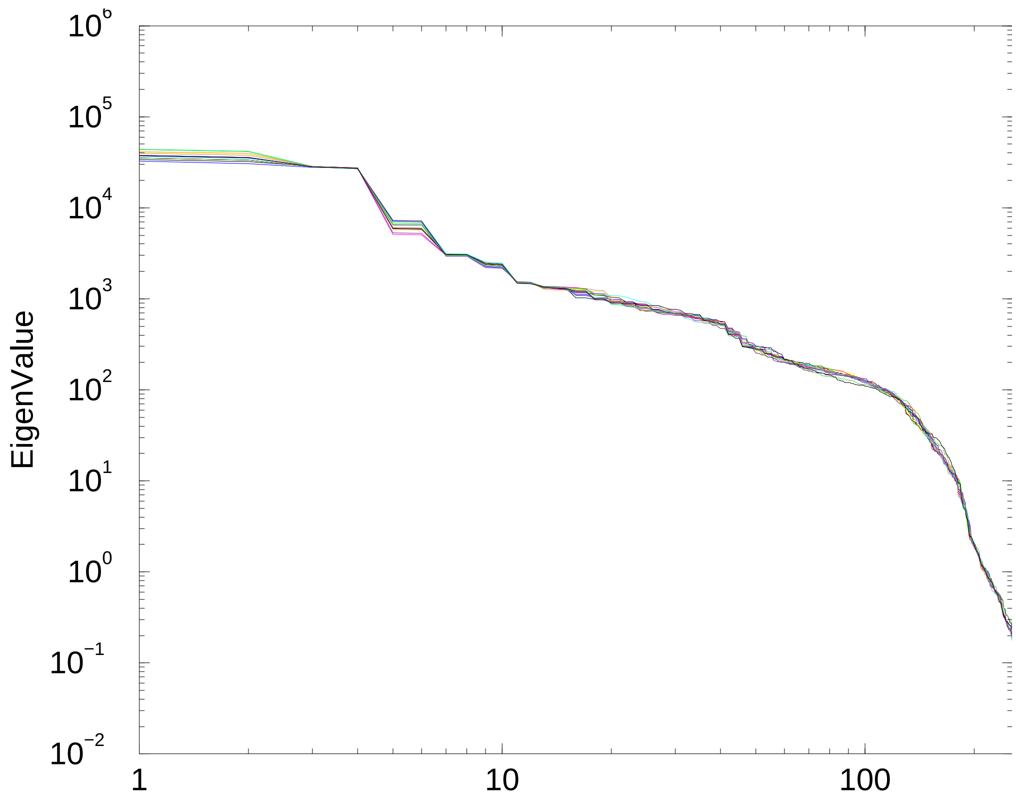
Notice that the DKL basis does not depend on the position of the analysis window (if the noise is *stationary*). Of course, the coefficients c_k do.

An example of c_k coefficients from LIGO 40m data



Hierarchy of the c_k coefficients (40m data)

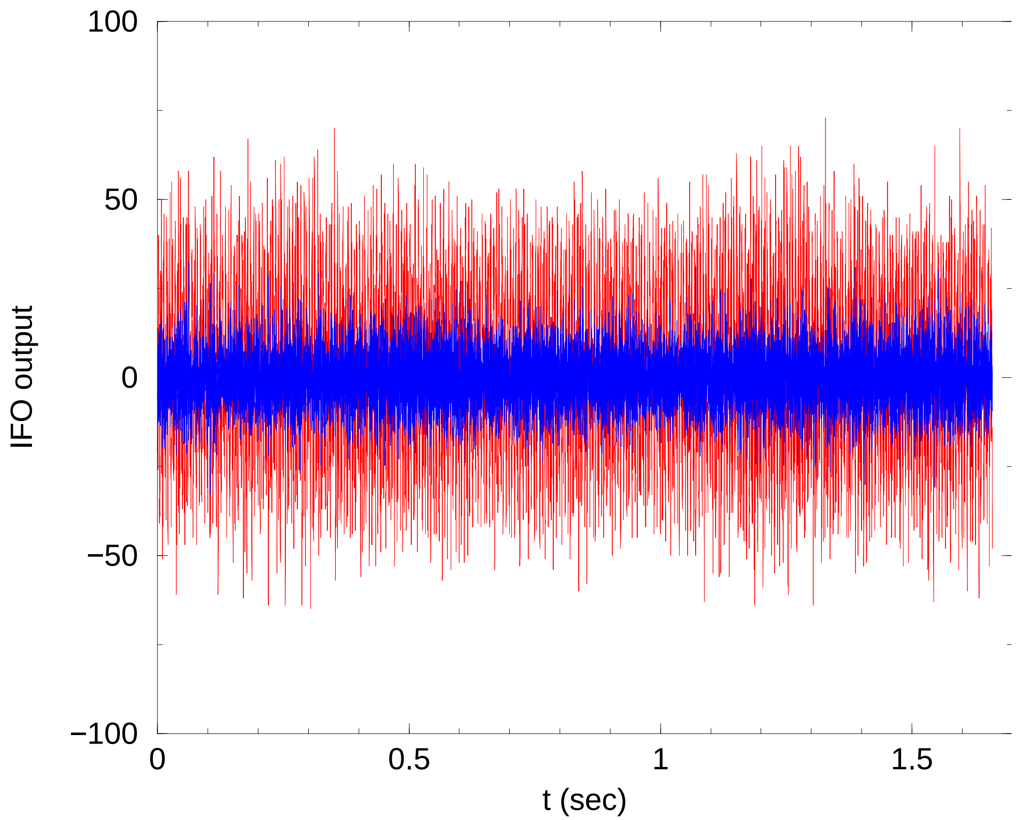
The variances σ_k of the c_k coefficients induce a natural order in the ψ^k basis



a larger variance corresponds to a “more important” or “noisy” component.

DKL variances and the RMS noise

It should be clear that Poisson theorem holds, and the RMS noise is just the sum of the variances σ_k corresponding to the DKL components



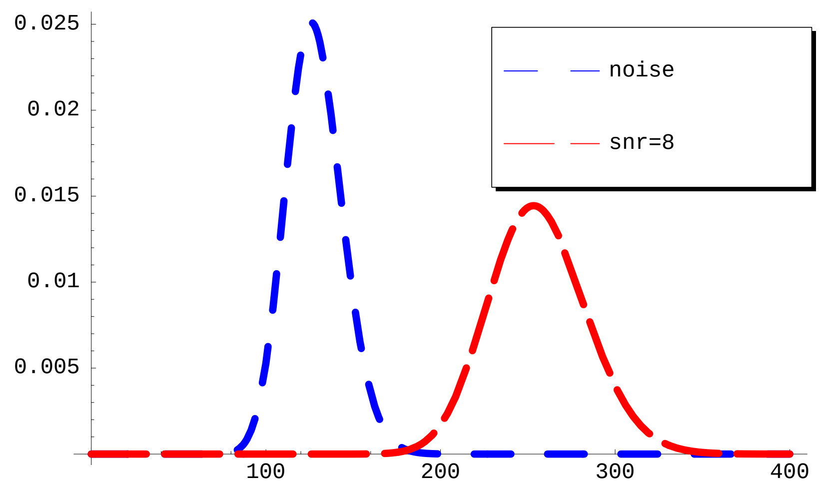
as can be visualized subtracting some of the more noisier components.

Statistics

It can be directly shown that each DKL coefficient is distributed as a Gaussian variable, hence the L statistic is a χ^2 with N_{burst} degrees of freedom

$$d\left(L|\mathrm{SNR}\right) = \frac{L^{N_{\mathrm{burst}}/2-1} e^{-\frac{1}{2}\left(L+\sqrt{2N_{\mathrm{burst}}}\,\mathrm{SNR}\right)}}{2^{N_{\mathrm{burst}}/2}\Gamma(N_{\mathrm{burst}}/2)}\, {}_0\mathrm{F}_1\left(;\frac{N_{\mathrm{burst}}}{2};\frac{\mathrm{SNR}\,L\sqrt{2N_{\mathrm{burst}}}}{4}\right);$$

the “pure noise” case is recovered with SNR=0



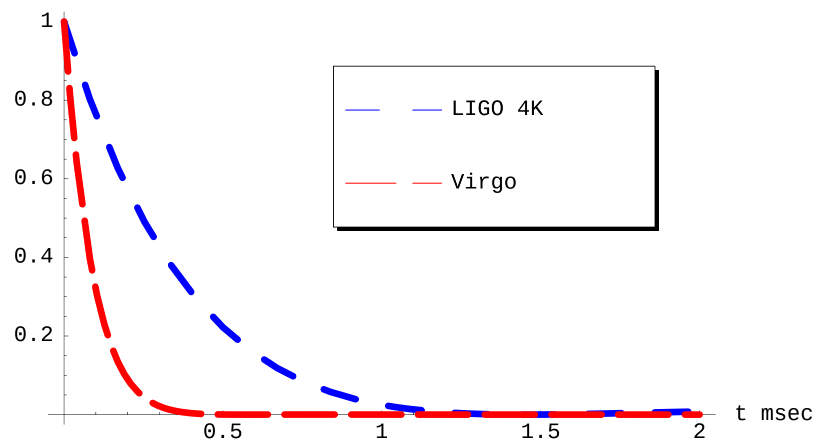
we need to assign a L value to $d\left(L|0\right)$ or $d\left(L|\mathrm{SNR}\right)$ for some SNR.

Outputs are cross-correlated

Instances of the statistic L relative to different time windows are correlated: for instance for $N_{\text{burst}} = 1$ one has simply

$$\frac{1}{2}\left(\frac{E\left[L_a\left(\mathbf{x}\right)L_b\left(\mathbf{x}\right)\right]}{E\left[L_a\left(\mathbf{x}\right)\right]E\left[L_b\left(\mathbf{x}\right)\right]} - 1\right) = \left[\frac{\left(\mathbf{R}_n^{-1}\right)_{ab}}{\left(\mathbf{R}_n^{-1}\right)_{aa}}\right]^2$$

and the correlation typically decays exponentially



which sets the sampling frequency for L.

A tool for event characterization

We have argued that the event detection corresponds to some thresholding on the L statistic: given that, a burst can be characterized by

✔ the noise background:

- ✘ the spectrum S_n at resolution $1/N$, needed for δ filtering
- ✘ and S_y at resolution $1/N_{\rm burst}$ (or alternatively $\mathbf{R}_y$) , needed to build the statistic.

✔ The event characteristics:

- ✘ the time window where the threshold is exceeded
- ✘ the coefficients of the DKL decomposition, in the $N_{\rm burst}$ subspace, which deviate from the expected Gaussian distribution.

In short: a Principal Component Analysis. In particular, only a subset of the DKL coefficients may be needed to keep a record of the event.

Extension to the network case

There are three more ingredients that we need in order to generalize the algorithm to a network of interferometric detectors

1. appropriate factors and delays are needed, depending on the direction of the source, which needs to be scanned.

2. Each detector comes with a different noise distribution, and

3. the noise can be correlated among the detectors.

Not too surprisingly, these ingredients contribute essentially only to the δ-filtering step.

The reason is that, given the chosen prior for the signal, the detection is truly a search for δ-events (coherent step) followed by a quadratic sum, properly weighted, over the filter outcomes (un-coherent step).

A network filter for δ-like events

The first step is to search for δ-events

$$y_{\mathsf{L}}[l] \equiv \sum_{\mathsf{K}k} \left(R^{-1}\right)_{\mathsf{LK}lk} x_{\mathsf{K}}[k+d_{\mathsf{K}}] \; ,$$

where R is a 4 index matrix, with indices running over detectors and lags.

Its approximate expression is much more intuitive

$$y_{\mathsf{L}}[l] \simeq \frac{2}{N} \sum_{k=1}^{N-2} \frac{e^{-i2\pi kl/N}}{S_{\mathsf{LL}}[k]} \left[\tilde{x}_{\mathsf{L}}[k] - \sum_{\mathsf{K} \neq \mathsf{L}} \frac{S_{\mathsf{LK}}[k]}{S_{\mathsf{KK}}[k]} \tilde{x}_{\mathsf{K}}[k] \right] \; ;$$

and hopefully adequate, as long as the cross-spectra S_{LK} are small with respect to the noise spectra S_{LL} in individual detectors.

Antenna patterns, and the "network" statistic

One needs to introduce the antenna patterns

$$\mathsf{F}_\mathsf{L}(\theta,\phi) \equiv \begin{pmatrix} \mathsf{F}_\mathsf{L}^{+} \\ \mathsf{F}_\mathsf{L}^{\times} \end{pmatrix}$$

and a $2 \times N$ matrix combining the data and the patterns

$$\mathsf{y} \equiv \begin{pmatrix} \mathsf{y}_{+} \\ \mathsf{y}_{\times} \end{pmatrix} = \sum_\mathsf{L} \mathsf{F}_\mathsf{L} \otimes \mathbf{y}_\mathsf{L};$$

it depends on the direction in the sky, but not very rapidly.

Now $\mathsf{y}_{\|}$ shall have a $(2 \times N_{\mathrm{burst}}) \times (2 \times N_{\mathrm{burst}})$ correlation matrix

$$\Theta = E\left[\mathsf{y}_{\|} \otimes \mathsf{y}_{\|}\right]$$

and

$$L = \mathsf{y}_{\|} \cdot \Theta^{-1} \cdot \mathsf{y}_{\|}$$

is the formal expression for the network statistic.

The devil hides in the details

The practical implementation requires to introduce two DKL bases

$$\psi_+^k \quad \text{and} \quad \psi_\times^k$$

for the two polarizations: but Θ^{-1} is complicated because $\psi_+^k \cdot \psi_\times^l \neq \delta_{kl}$.

Worse, both bases depend on the direction in the sky!

Only if $N_{\rm burst}$ is large, the DFT can replace the DKL and things get simpler: from $\mathsf{y}_+, \mathsf{y}_\times$ we compute the "network" spectrum

$$\mathsf{S}_y[k] \equiv \begin{pmatrix} S_{++}[k] & S_{+\times}[k] \\ S_{\times+}[k] & S_{\times\times}[k] \end{pmatrix}$$

and the statistic becomes

$$L \propto \sum_{k=1}^{N_{\rm burst}-2} \left[\tilde{\mathsf{y}}[k]\right]_{\|}^H \cdot \left[\mathsf{S}_{\mathsf{y}}[k]\right]^{-1} \cdot \left[\tilde{\mathsf{y}}[k]\right]_{\|}$$

which is however just an approximation.

A simple example: a network search for δ events

Suppose we use the global GW network to search for events affecting just one sample: then

$$\tilde{x}_{\mathrm{L}}[k] = \left(A_+ F_{\mathrm{L}}^+ + A_\times F_L^\times\right) \frac{1}{f_s} e^{-j2\pi ka/N}$$

is the response of each detector. The matrix Θ becomes simply

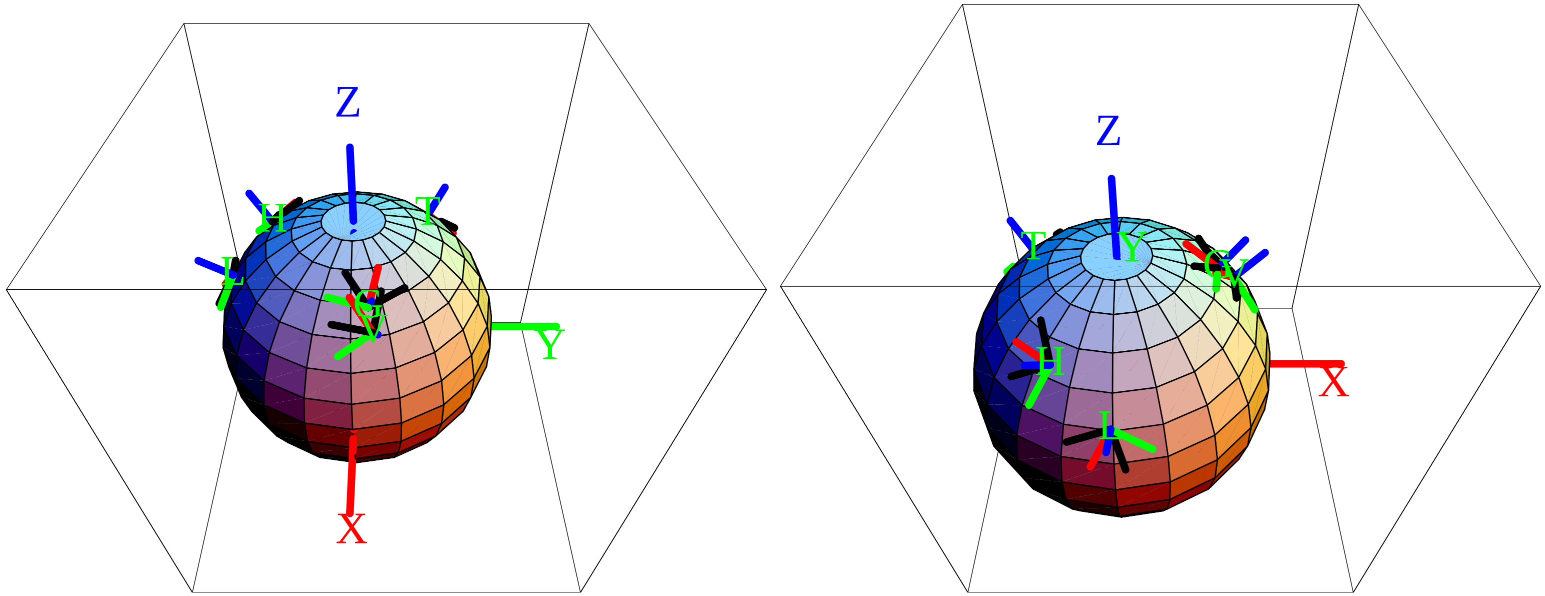
$$\Theta = \sum_{\mathrm{L}} \mathrm{rms}\left(\mathrm{y_L}\right) \begin{pmatrix} \left(F_{\mathrm{L}}^+\right)^2 & F_{\mathrm{L}}^+ F_{\mathrm{L}}^\times \\ F_{\mathrm{L}}^+ F_{\mathrm{L}}^\times & \left(F_{\mathrm{L}}^\times\right)^2 \end{pmatrix}$$

where y_{L} is computed as before, and in absence of noise is

$$\mathrm{y}_{\parallel} = \sum_{\mathrm{L}} \mathrm{rms}\left(y_{\mathrm{L}}\right) \left(A_+ F_{\mathrm{L}}^+ + A_\times F_{\mathrm{L}}^\times\right) \begin{pmatrix} F_{\mathrm{L}}^+ \\ F_{\mathrm{L}}^\times \end{pmatrix};$$

we can assemble the statistic $L = \mathrm{y}_{\parallel} \cdot \Theta^{-1} \cdot \mathrm{y}_{\parallel}$.

The global network



Black lines represent the ITF axes.

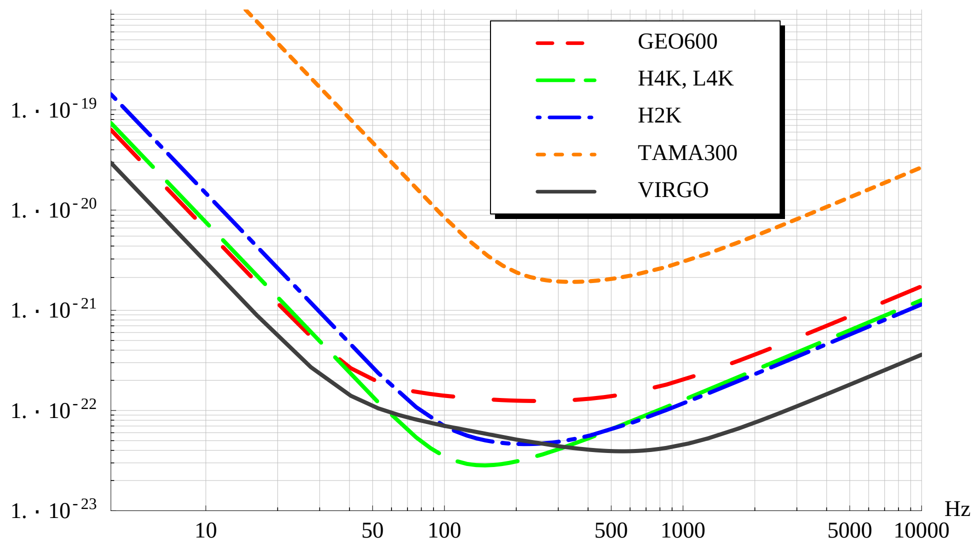
Colored lines represent the axes of the detector frames and of the Earth frame.

Z crosses the North pole

X crosses the Greenwich meridian

Noise in the individual detectors

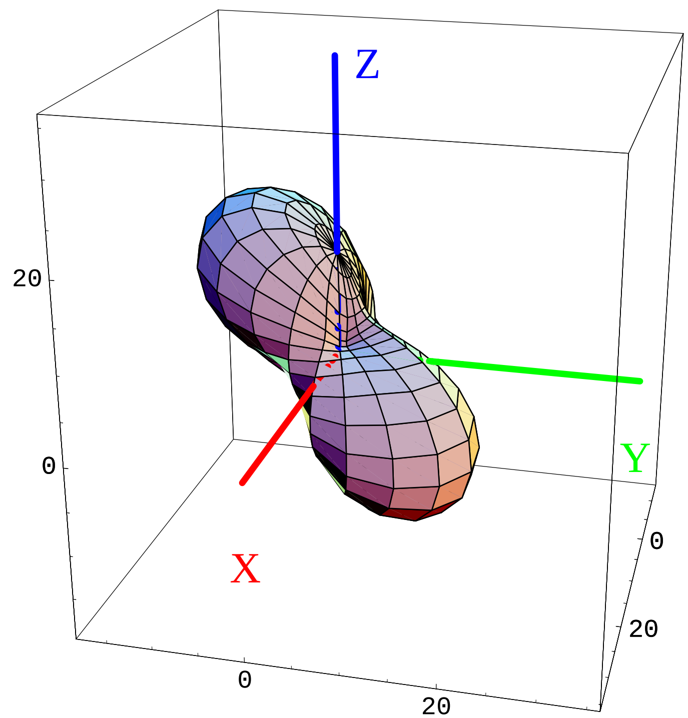
The design sensitivities



are needed only to compute

$$\mathrm{rms}\,(y_{\mathsf{L}}) \equiv \frac{1}{f_s N} \sum_{k=1}^{N/2-1} \frac{1}{S_{\mathsf{LL}}[k]} \simeq \frac{1}{f_s^2} \int_{f_{\mathsf{seism}}}^{f_{\mathsf{Nyquist}}} \frac{df}{S_{\mathsf{LL}}(f)}.$$

The response of the LIGO network

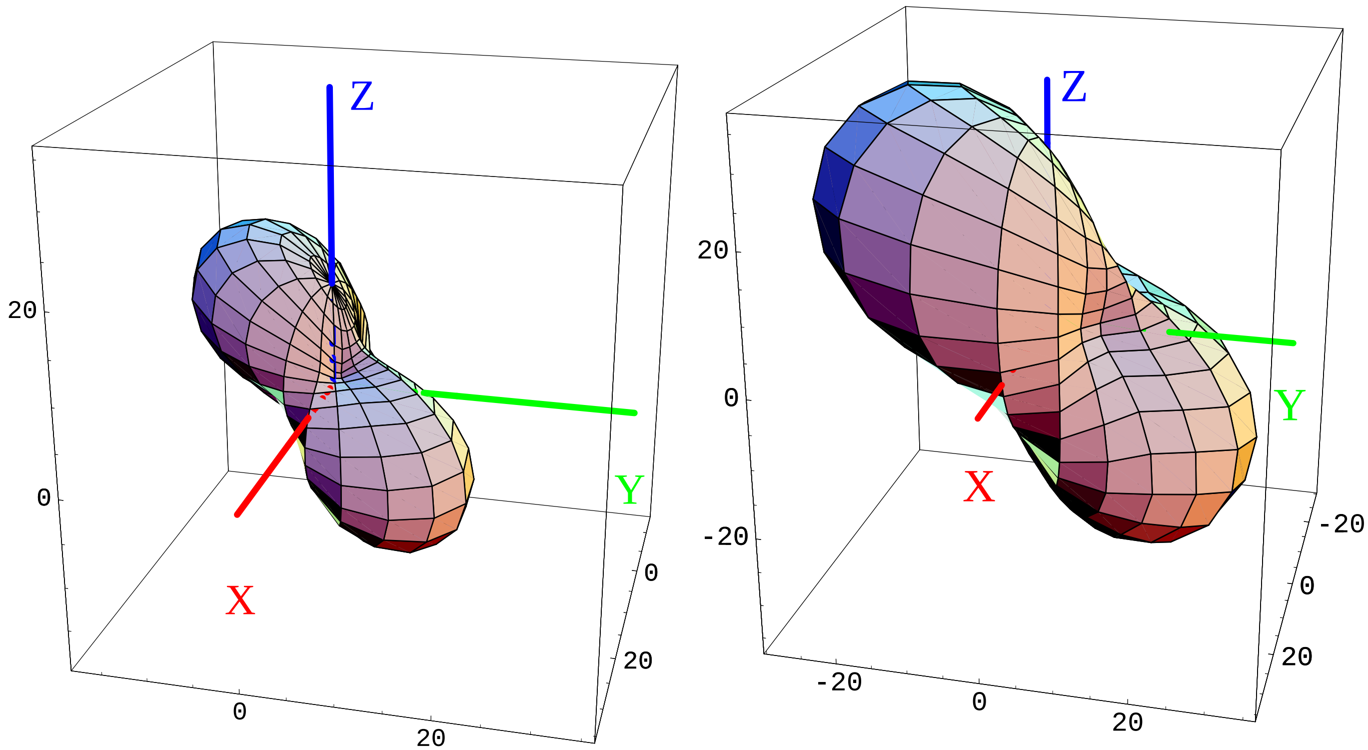


An average was done on the signal polarization, keeping

$$A \equiv \sqrt{A_+^2 + A_\times^2}$$

fixed. Numerically $A\,dt = 10^{-23} s$, corresponding to $A = 10^{-20}$ for a 1ms burst.

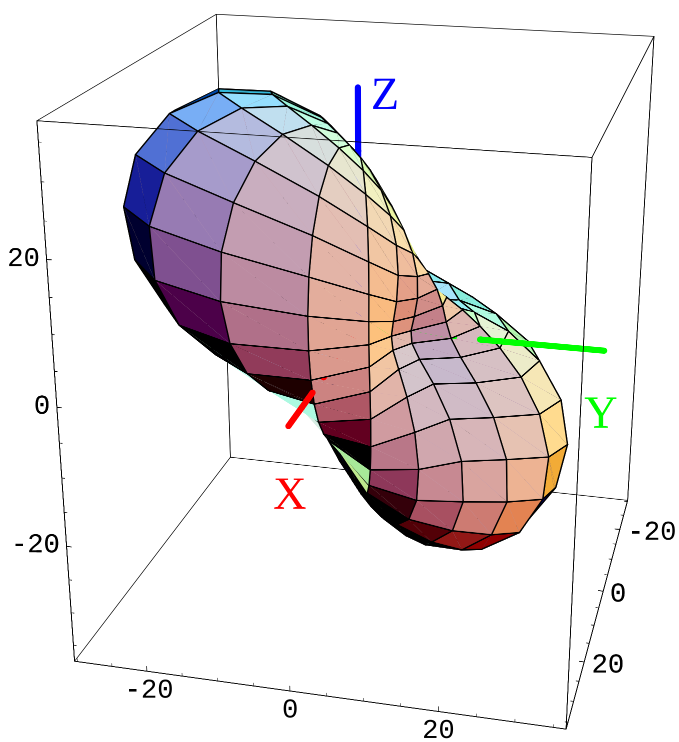
Adding Virgo to the network

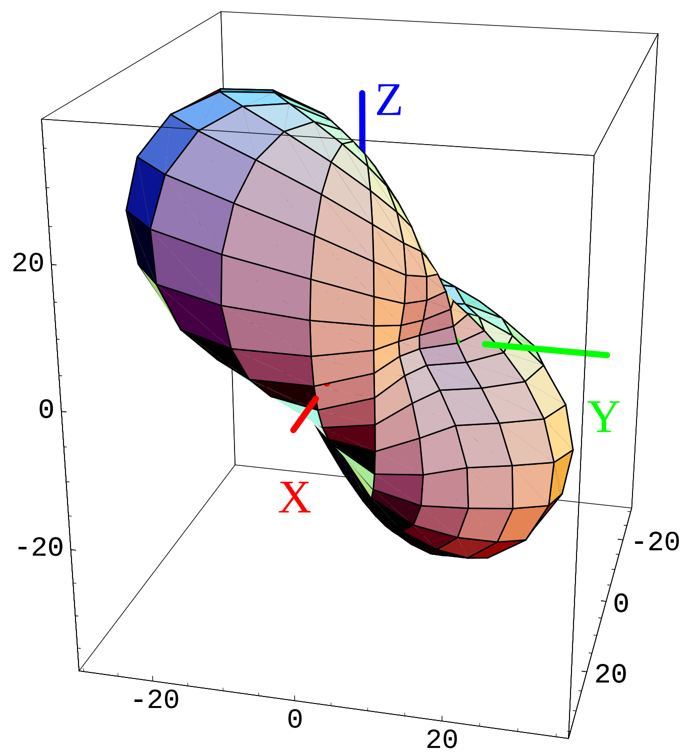


Left: LIGO alone, right LIGO+Virgo.

The Virgo detector gives a substantial enhancement thanks to its wider bandwidth, despite the lower sensitivity in the $60 - 300$ Hz band

Adding GEO and TAMA

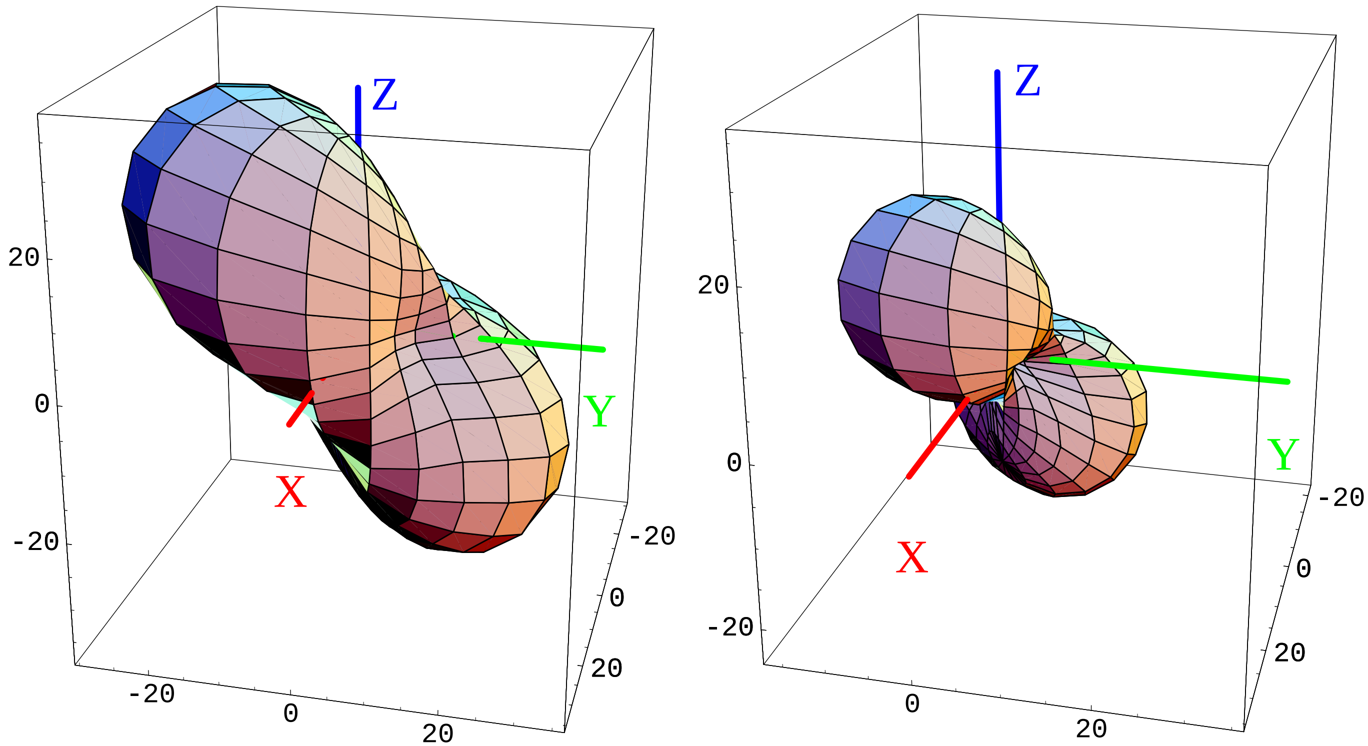




Left: LIGO + Virgo. Right: including GEO and TAMA.

With the current *design* sensitivity, they don't make a great difference.

Whole network versus GEO and Virgo alone



Left: all the ITFs. Right: just GEO and Virgo.

The statistic L is quadratic in the signal amplitude, thus enhancing the visual effect when combining detectors. Of course the detection probability matters, and should be gauged on the χ^2 distribution.

Conclusions

✔ An algorithm for burst detection "optimal" under the assumption of total ignorance about the signal waveform was deduced.

✔ It can be exploited both to detect and to characterize an event. A true signal, or a burst of non-gaussian noise.

✔ The single detector form is simple: the multi-detector form is much more complicated, apart the two limiting cases of $N_{\rm burst} = 1$ or $N_{\rm burst} \ll 1$.

✔ If (as it is very likely) there is significant non-gaussian noise in the detectors, probably a coincidence strategy is more effective.

✔ A fuller description can be found in the preprint LIGO-P010019-00-E [7].

✔ People willing to create plots of the network sensitivity (for δ-bursts) varying viewpoints, detector locations, sensitivities, can download a *Mathematica* notebook at http://www.ligo.caltech.edu/ avicere/nda/burst/Burst.nb

References

[1] T.Zwerger and E.Müller, Astron. Astrophys. **267**, 623 (1993).

[2] N.Arnaud, M.Davier, F.Cavalier, P.Hello, Phys.Rev.D **59**, 082002 (1999).

[3] T.Pradier, N.Arnaud, M.A.Bizouard, F.Cavalier, M.Davier, P.Hello, Int. J. Mod. Phys. D **9**, 309-314 (2000).

[4] T.Pradier, N.Arnaud, M.A.Bizouard, F.Cavalier, M.Davier, P.Hello, Phys. Rev. D **63,** 042002 (2001).

[5] É.É.Flanagan, S.A.Hughes, Phys. Rev. D **57**, 4566 (1998).

[6] W.G.Anderson, P.R.Brady, J.D.E.Creighton, É.É.Flanagan, Phys. Rev. D **63**, 042003 (2001).

[7] A.Viceré, *Optimal detection of burst events in gravitational wave interferometric observatories*, LIGO-P010019-00-E.